THE AMERICAN MATHEMATICAL MONTHLY

Volume 93, Number 1 January 1986

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. (This is necessary for indexing purposes.) Send all main articles, letters to the editor, and editorial correspondence to the editor. Contributions to the other four special departments (Notes, Teaching of Mathematics, Unsolved Problems, and Problems and Solutions) should be sent to the appropriate one of the associate editors, whose addresses are given below.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT, ten times a year: January, February, March, April, May, June-July, August-September, October, November and December.

The annual subscription price for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is $24 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $24. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $70 per year.

Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

PRINTED IN THE UNITED STATES OF AMERICA

THE AMERICAN
MATHEMATICAL MONTHLY

AN OFFICIAL JOURNAL OF

THE MATHEMATICAL ASSOCIATION OF AMERICA

(INCORPORATED)

DEVOTED TO THE INTERESTS OF COLLEGIATE MATHEMATICS

THE AMERICAN MATHEMATICAL MONTHLY, FOUNDED IN 1894 BY BENJAMIN F. FINKEL, WAS PUBLISHED BY HIM UNTIL 1913. FROM 1913 TO 1916 IT WAS OWNED AND PUBLISHED BY REPRESENTATIVES OF FOURTEEN UNIVERSITIES AND COLLEGES IN THE MIDDLE WEST

VOLUME 93

1986

PUBLISHED BY THE ASSOCIATION
MONTPELIER, VT, AND WASHINGTON, DC

LET'S MEET AT THE CONGRESS

YOUSEF ALAVI
Department of Mathematics, Western Michigan University, Kalamazoo, MI 49008

PETER HILTON
Mathematical Sciences Department, SUNY at Binghamton, Binghamton, NY 13901

JEAN PEDERSEN
Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053

for the Public Information Committee of ICM–86*

1. What is ICM? Do you know what the initials ICM stand for? The answer is *not* the International Congress of *Mathematics*—it is the International Congress of *Mathematicians*. An ICM is a gathering of people with a common interest, mathematics; it is not merely, or even primarily, the opportunity to report on the latest results of mathematical research. The scientific program is designed to be a synthesis of the developments that have taken place in mathematics in the past four years, and, as such, it presents features of interest for each and every one of us; and the social program is designed to provide opportunity for mathematicians from all over the world to meet together to discuss their discipline, their field, their art. Organized under the auspices of the International Mathematical Union, the Congress gathers together thousands of mathematicians and gives them a chance to talk about the latest exciting discoveries and the latest professional gossip. An ICM can thus exercise a profound influence on the direction of mathematics in all its various aspects.

A highlight of the Congress is the award of Fields Medals to young mathematicians to recognize their outstanding achievements. Sometimes two medalists are named, sometimes as many as four. Lars Ahlfors and Jesse Douglas were the first two winners (in 1936), and at the Warsaw Congress (in 1982) it went three ways: Alain Connes, William Thurston, and Shing-Tung Yau. Many of the winners in the years between have become household names in the family of mathematicians—they include Laurent Schwartz, Klaus Roth, Lars Hörmander, John Milnor, Michael Atiyah, Paul Cohen, Enrico Bombieri, and Gregori Margulis—and that's only about half of them. There is no Nobel Prize in mathematics (there are many explanations of this extraordinary omission on Alfred Nobel's part, but none of the more titillating versions has been authenticated!), and we recognize the Fields Medal as the highest honor we can pay to a research mathematician. Those of us who attended the 1954 Congress in Amsterdam will remember it, among other things, as the first occasion on which Jean-Pierre Serre was seen wearing a tie—what better testimony to the importance that attaches to the award! At Amsterdam and Stockholm, Dutch and Swedish royalty graced the occasion of the Fields Medals awards—we cannot promise a comparable recognition in Berkeley, but at least we ourselves will give due appreciation to the great work of the recipients of the award, and thereby signify our respect for our science. It should be kept in mind, however, that the Fields Medal and the Nobel Prize differ in at least two ways: in money and in age. You don't get rich when you are awarded a Fields medal, and you have to be young—less than or equal to 40.

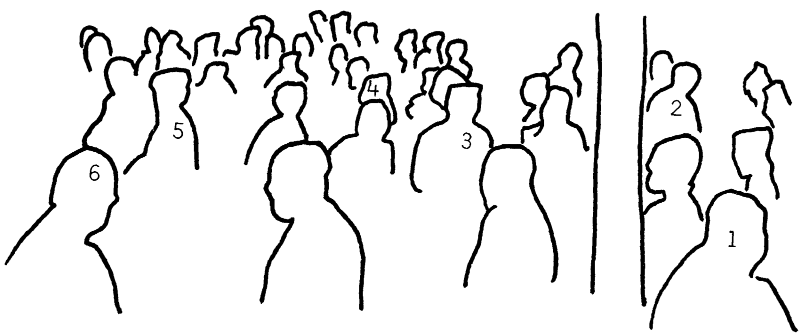
This is only the third time since the first Congress in 1893 that an ICM is being held in the United States, the others having been the 1893 Congress at Northwestern University and the 1950 Congress at Harvard. ICM–86, which will be held at the University of California at Berkeley from Sunday, August 3, through Monday, August 11, 1986, is not only a great opportunity for

*The committee takes responsibility for the facts contained in this article, but firmly assigns to the authors responsibility for the selection of facts and the statement of opinions.

mathematicians to come together but also a great opportunity to explain to those yet to be convinced the importance of mathematics itself. It is a fact, especially conspicuous in the United States, that the importance of mathematics, indeed the very nature of mathematics, is not well understood by politicians and policy-makers. There are two particularly erroneous views to be combatted. First we must convince the public that mathematics is not the same as computation, and second that it is not true that all that is really important in mathematics was known long ago. Of course, mathematicians know this—but many people outside mathematics do not. One of the principal tasks of the committee we represent is to tell the world how lively and vigorous mathematics is today.

2. History of the ICM's. The ICM's have been held regularly at four-year intervals since 1893 except for the years of the two World Wars. An exception was made in 1900, a year which mathematicians found irresistible for the holding of a Congress. That was the celebrated Paris Congress at which David Hilbert presented the famous 23 open problems of mathematics. From 1900 the congresses were held every year $n \equiv 0$ modulo 4 through 1936, except for 1916. (It may be necessary to find a less elegant formulation of this modular pattern for the benefit of math avoiders, including perhaps some journalists and other luminaries of the media.) The 1936 Oslo Congress was the last before the Second World War; Harvard University had issued an invitation for 1940 but that Congress had to be postponed until 1950. Thereafter congresses were held every year $n \equiv 2$ modulo 4 up to the Helsinki Congress of 1978.

This picture was taken at the Bologna Congress in 1928. It contains Ingham, (unknown), Pólya, Szegő, Mrs. Reidemeister, Reidemeister, Doetsch.



This group picture was taken at the Oslo Congress. It includes some well-known faces: (1) Wiener, (2) Weyl, (3) Fréchet, (4) Borel, (5) Veblen, and (6) Carathéodory.

The complete list of these Congresses reads as follows:

1954	Amsterdam	(Holland)
1958	Edinburgh	(Scotland)
1962	Stockholm	(Sweden)
1966	Moscow	(USSR)
1970	Nice	(France)
1974	Vancouver	(Canada)
1978	Helsinki	(Finland)

The 1966 Congress was remarkable in many ways. Naturally, the attendance of Soviet mathematicians was far greater than at any other Congress (about 80 times as great, in fact—Sammy

Eilenberg, at the Congress, toasted the population explosion that had evidently taken place in the USSR!). But this was also the occasion on which Steve Smale gave his celebrated press conference on the steps of Moscow University—for further details consult his forthcoming autobiography.

The year 1982 was to have seen the Warsaw Congress. Since, however, martial law was introduced in Poland in December, 1981, and was still in force in the summer of 1982, it became obvious that suitable conditions did not exist for the holding of an international scientific gathering. The IMU decided that the Warsaw Congress should either be postponed to the summer of 1983, or cancelled, and that a decision on which of these alternatives to adopt should be taken following a 'site visit' in the fall of 1982 by members of the Executive Committee of IMU. The visiting committee received assurances from a Polish Government spokesman that there would be no adverse circumstances attending the holding of a congress in the summer of 1983, and it was consequently decided to go ahead, with the original scientific program. Not everybody agreed with this decision; moreover, American attendance was substantially reduced by the refusal of the National Science Foundation to grant any travel funds to support participation at the Congress.

The Warsaw Congress, scheduled for 1982, was, as we have said, postponed until 1983, but the modular pattern is being restored by holding the next Congress in Berkeley in the summer of 1986, at the invitation of the National Academy of Sciences.

The attendances at the ICM's have been growing steadily (but not quite monotonically) since their inception, from a few hundred to several thousand. It is expected that there will be around 4,000 participants at ICM–86. The scientific scope of the ICM's has also increased very noticeably. Thus at Harvard in 1950, mathematics was divided into 7 Sections, whereas at Berkeley there will be 19 Sections. It has also become customary to hold smaller, specialized meetings at times and places which make it convenient for Congress participants to attend, and we have every reason to suppose that this pattern will be followed in connection with the Berkeley Congress.

3. Structure of ICM–86. Traditionally, the International Congress is run by an Organizing Committee of volunteer scientists at and around the site university. ICM–86 is a break with that tradition. The National Academy of Sciences asked the American Mathematical Society to handle the organizational aspects of the Congress, and the Society has chosen to do so through a newly-formed subsidiary corporation. There is a Steering Committee of twelve mathematicians representing many of the mathematical organizations in the United States. The office of the congress manager handles the details (travel, housing, meeting rooms, entertainment, and many others) with advice from the Steering Committee and occasional consultation with the mathematicians of the Bay Area near Berkeley.

Funds to support ICM–86 will be obtained from three sources: registration fees from the participants, government grants, and contributions from the private sector. It follows as a corollary that the registration fee will be reduced in proportion to the extent of success in obtaining funds from the other two sources named. It is hoped to keep the registration fee down to 833% of the 1950 figure!

The responsibility for the scientific program lies with the International Mathematical Union through its designated Consultative Committee. This committee consists of world class mathematicians drawn from the international community. The members of the committee function as individuals, expert in their fields, during the long drawn out process of developing the scientific program. It should especially be emphasized that they base their deliberations exclusively on criteria relevant to the merit of the program and not on national or political considerations. Indeed, this has occasionally caused problems, as, for example, when Pontryagin declared, on behalf of the Soviet National Committee for Mathematics, that he was not in complete agreement (to put the matter mildly!) with the list of invitations issued to Soviet mathematicians to speak at the Helsinki Congress.

At ICM–86 the major areas of mathematics will be covered by 17 invited one-hour plenary addresses, intended to be accessible to all mathematicians. There will also be about 130 invited

lectures, of 45 minutes duration, in the 19 Sections into which mathematics is divided at the Berkeley Congress. Congress participants are divided into *Ordinary Members*, that is, those who plan to attend the scientific sessions and to receive the proceedings; and *Accompanying Members*, that is, those wishing only to participate in extra-mathematical activities. The overall program of the Congress includes a number of social and cultural events, and, in particular, the Chancellor of the University of California at Berkeley will host a reception for *all* Congress members.

The official languages of the Congress are English, French, German and Russian; one may reasonably anticipate that the great majority of the talks will be in English. It is interesting, incidentally, to notice that Italian was also an official language of the Harvard Congress of 1950; and that Spanish is not an official language of ICM–86, although it is the first language of many mathematicians (in particular, all Latin American mathematicians, except the Brazilians). Fortunately for American participants, there is unlikely to be a very severe language barrier, since practically all mathematicians the whole world over have become proficient in the English language, in conversation as well as in mathematical exposition.

4. The Benefits of Attendance. Undoubtedly an ICM is an event of great significance in the evolving history of our subject. The individual mathematician, however, is bound to ask—what are the benefits to me of attending? After all, the invited addresses are published. Moreover, there are clear disadvantages in attending—there is the cost, and, probably, there would be involved some disruption of family life during the summer vacation period. What, then, are the advantages to be set against these arguments?

First and foremost there are the professional advantages and, among these, we may give pride of place to the opportunity to meet our great contemporaries. Every mathematician should feel some excitement when the citations for the Fields Medalists are read out, even if not all details of their contributions are understood. These are our "Nobel prizewinners", and as we meet them and listen to them we must get a marvelous sense of participating, however modestly, in a great and continuing achievement of the human spirit.

We have the opportunity to listen to interesting lectures delivered by famous research mathematicians from many countries and thereby to bring ourselves more up-to-date in the state of the art. The program, as we have said, includes 17 invited lectures at plenary sessions intended to be accessible to all Congress participants. There are also the 130 invited lectures distributed over 19 Sections, or special areas, and addressed to those who have advanced knowledge in those areas. We should emphasize that, at this Congress, special provision is being made for the interests of those who are concerned not only with research but also with the necessary adaptations of the curriculum, as a consequence of the new uses of mathematics. Sections 14 through 19 are entitled Numerical Methods and Computing; Discrete Mathematics and Combinatorics; Mathematical Aspects of Computer Science; Applications of Mathematics to Non-physical Sciences; History of Mathematics; Teaching of Mathematics.

We may also avail ourselves of the chance to present a 10-minute communication ourselves, thus letting the mathematical world know what we are doing and enabling those attending the Congress to identify our field of interest, to the subsequent benefit of our own research. It is anticipated that room will be found in the program for some 1000 such communications, a number considerably in excess of that at previous congresses. However, those contemplating presenting such a communication should be warned that 10 minutes will seem much too short a time in which to say something of significance in mathematics, and that those presiding at the sessions for short communications will have been instructed to be firm with speakers who attempt to exceed their allotted time! It is much easier to design a 50-minute talk than a 10-minute talk!

There are reasons for attending the Congress that have both professional and social aspects. It is interesting, and good fun, to meet other mathematicians, particularly those from other countries, and to get a real sense of belonging to a larger community. It is rewarding to discuss common problems of teaching mathematics with mathematicians from different social and political backgrounds, and sometimes comforting to find how much our problems have in

common—and how far we all are from solutions! But it is also pleasant simply to relax among those who share our view of the importance of mathematics. The social events are designed to give the maximum opportunity for informal, unrestricted contact.

Probably for many participants this may be the most vivid impression they carry away with them from the Congress—this sense of belonging to a community which transcends all barriers of nationhood, of politics, of race and religion—this sense that a common love of mathematics is a stronger force in uniting our profession than any other force could be which seeks to pull us apart. And we do not believe that there is any better way to create this wonderful sense of genuine comradeship than participation in an International Congress.

A CONJECTURED ANALOGUE OF ROLLE'S THEOREM FOR POLYNOMIALS WITH REAL OR COMPLEX COEFFICIENTS

I. J. SCHOENBERG
Mathematics Research Center, University of Wisconsin, Madison, WI 53705

1. Introduction. We assume that $n \geqslant 2$ and let

$$P_n(z) = z^n + a_2 z^{n-2} + \cdots + a_n = \prod_1^n (z - z_j) \tag{1.1}$$

be a polynomial with $a_1 = 0$; hence its zeros z_j satisfy

$$z_1 + z_2 + \cdots + z_n = 0. \tag{1.2}$$

We also consider its derivative

$$P_n'(z) = nz^{n-1} + (n-2)a_2 z^{n-3} + \cdots + a_{n-1} = n\prod_1^{n-1} (z - w_k), \tag{1.3}$$

having the zeros w_k satisfying

$$w_1 + w_2 + \cdots + w_{n-1} = 0. \tag{1.4}$$

Evidently, the origin O of the complex plane $\mathbb{C}$ is the centroid of the set z_j and also the centroid of the w_k.

The set of $2n - 1$ points

$$R = \{z_j; w_k\} \tag{1.5}$$

we call *the complex Rolle set of* $P_n(z)$. If we turn the set R about the origin O by the angle θ, *the new set* $R(\theta)$ *is also a complex Rolle set.*

Indeed, by (1.1) and (1.3) we have

$$P_n(ze^{-i\theta}) = \prod_1^n \left(ze^{-i\theta} - z_j\right) = e^{-in\theta}\prod_1^n \left(z - z_j e^{i\theta}\right),$$

$$(d/dz)P_n(ze^{-i\theta}) = e^{-i\theta}P_n'(ze^{-i\theta}) = ne^{-i\theta}\prod_1^{n-1}\left(ze^{-i\theta} - w_k\right)$$

I. J. Schoenberg: I would very much like to know if our easy derivation of Theorem 2 from van den Berg's Theorem 1 generalizes so that Conjecture 1 can be derived from Theorem (4.2), on page 11, of Morris Marden's book [**2**, Theorem (4.2), page 11, for $p = n$ and $m_1 = m_2 = \cdots = m_n = 1$]. The w_k appear there as the foci of an algebraic curve of class $n - 1$ which touches all segments $z_r z_s$ $(r \neq s)$ in their midpoints. A short biography of mine appears on the back cover of the paper edition of my little book [**3**]. I may add to it that I played the violin and made chamber music most of my life, and that I have been a vegetarian since the age of ten.

$$= ne^{-in\theta}\prod_{1}^{n-1}\left(z - w_k e^{i\theta}\right).$$

These equations show that *the rotated set*

$$R(\theta) = \left\{ z_j e^{i\theta}, w_k e^{i\theta} \right\} \tag{1.6}$$

is the complex Rolle set of $P_n(ze^{-i\theta})$.

We say that the Rolle set R is *rectilinear*, provided that all its $2n-1$ points z_j and w_k are on a straight line which must contain the centroid O of all these points. If all z_j are real, then by the usual Rolle theorem also all w_k are real. The invariance of the Rolle set by rotation implies that if all z_j are on a line L through O, then all w_k are on L, and so R is rectilinear.

From (1.2) and (1.4) we obtain the equations

$$\sum_{1}^{n} z_j^2 = -2\sum_{j<j'} z_j z_{j'} = -2a_2, \tag{1.7}$$

$$w_k^2 = -\sum_{k<k'} w_k w_{k'} = -2\frac{n-2}{n}a_2, \tag{1.8}$$

whence, by eliminating a_2, we obtain the identity

$$\sum_{1}^{n-1} w_k^2 = \frac{n-2}{n}\sum_{1}^{n} z_j^2. \tag{1.9}$$

Thus the $2n-1$ points of any complex Rolle set R satisfy the important identity (1.9).

The invariance of R by rotation around O has the following consequence:

THEOREM 0. *If the Rolle set* $R = \{z_j; w_k\}$ *is rectilinear, then we have the equation*

$$\sum_{1}^{n-1} |w_k|^2 = \frac{n-2}{n}\sum_{1}^{n} |z_j|^2. \tag{1.10}$$

Indeed, the rectilinearity of R implies that there is an angle ϕ such that

$$z_j = \pm|z_j|e^{i\phi} \text{ for all } j, \quad \text{and} \quad w_k = \pm|w_k|e^{i\phi} \text{ for all } k.$$

Substituting these into (1.9), and canceling the factor $e^{2i\phi}$, we obtain the equation (1.10). We may therefore state

COROLLARY 1. *The equation* (1.10) *is a necessary condition for the rectilinearity of the Rolle set* $R = \{z_j; w_k\}$.

These results suggest the following

CONJECTURE 1. *For any complex Rolle set* $R = \{z_j; w_k\}$ *we have the inequality*

$$|w_1|^2 + \cdots + |w_{n-1}|^2 \leqslant \frac{n-2}{n}\left(|z_1|^2 + \cdots + |z_n|^2\right), \tag{1.11}$$

with the equality sign if and only if R is rectilinear.

In the sequel we first establish Conjecture 1 for the special case when $n = 3$; we actually settle this case in two different ways. For $n = 3$ the two points w_1 and w_2 are evidently on a line L through O. In Theorem 3 we prove Conjecture 1 for *any* complex Rolle set such that *all* w_k points are on a line L. In spite of the simple proof of Theorem 3 we have included the more complicated direct proof of Conjecture 1 for the case $n = 3$, because its approach might possibly generalize for arbitrary n (see the generalization for all n of our Theorem 1 to Theorem (2, 3) in Morris Marden's book [**2**, p. 11]. We also establish Conjecture 1 for the special case when $P_n(z)$ is a *binomial polynomial.*

Fred Sauer, of the MRC Computing Staff, has verified the inequality (1.11) for some 25 numerically given complex $P_n(z)$, for $n = 4$ and $n = 5$. For this help I am much obliged to Fred. I am also impressed by the speed and precision of the Jenkins-Traub algorithm used in solving the equations $P_n(z) = 0$ and $P_n'(z) = 0$.

2. A first proof of Conjecture 1 for $n = 3$. Let

$$T = (z_1, z_2, z_3) \tag{2.1}$$

be a non-degenerate triangle in the complex plane having the zeros of $P_3(z)$ as vertices. By (1.2) the centroid of T is in the origin O. We shall use the following theorem of van den Berg ([**1**], or [**3**, Chapter 7]).

THEOREM 1 (van den Berg). *Let E be the Steiner ellipse of the triangle. This is the ellipse which is inscribed in T such that E is tangent to the sides of T in their midpoints. Then the zeros of $P_3'(z)$ are identical with the focii w_1 and w_2 of the ellipse E* (see Fig. 1).

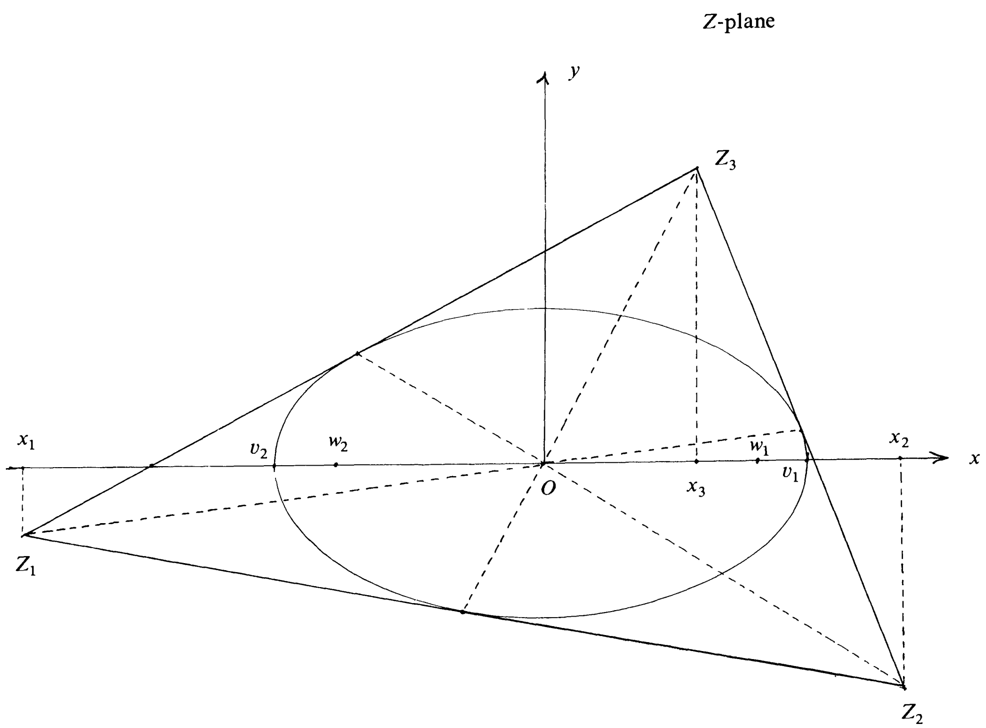


FIG. 1

With $z = x + iy$ and $z_j = x_j + iy_j$ we place $T \cup E$ so that the major axis v_1v_2 of E is on the real axis, its center being at the origin. Let a and b be the semi-axes of E and $a^2 - b^2 = c^2 = w_1^2 = w_2^2$. We now subject the plane to the affine transformation

$$A_t: \begin{array}{l} x(t) = x \\ y(t) = yt, \quad (0 \leqslant t \leqslant 1), \end{array}$$

which contracts E toward the x-axis. As the semi-axes of the new ellipse $E(t) = A_t E$ are a and bt, we find that the foci of $E(t)$, which we denote by $w_1(t)$ and $w_2(t)$, have the abscissae

$$w_1(t) = \sqrt{a^2 - b^2t^2} \quad \text{and} \quad w_2(t) = -\sqrt{a^2 - b^2t^2}.$$

As $t \to 0+$ we see

1. That the foci of $E(t)$ converge to the endpoints v_1 and v_2 of the major axis of E.
2. That $\lim z_j = x_j$ $(j = 1,2,3)$.

By continuity we conclude that $(x_1, x_2, x_3, v_1, v_2)$ is a complex Rolle set composed of five real points, so that by (1.9) we have $v_1^2 + v_2^2 = (1/3)(x_1^2 + x_2^2 + x_3^2)$ and therefore

$$\frac{1}{3}\sum_1^3 |z_j|^2 \geq \frac{1}{3}\sum_1^3 x_j^2 = v_1^2 + v_2^2 \geq |w_1|^2 + |w_2|^2. \tag{2.2}$$

A comparison of the extreme members proves the inequality (1.11).

Moreover, if the extreme members of (2.2) are equal, then we must have $\Sigma_1^3|z_j|^2 = \Sigma_1^3 x_j^2$. This clearly implies that $|z_j|^2 = x_j^2$ for all j, and therefore that $z_j = x_j$ for all j. We have therefore established

THEOREM 2. *Conjecture* 1 *holds for* $n = 3$.

3. A generalization. We shall now generalize Theorem 2 in a certain direction and at the same time simplify its proof, as Theorem 1 will not be used. However, we add the new assumption that

$$\textit{All } w_j \textit{ are on a line } L \textit{ through } O. \tag{3.1}$$

We may as well assume that all w_j are real. This allows us to state

THEOREM 3. *If all* w_j *are real, then we have the inequality*

$$\sum_1^{n-1} w_k^2 \leq \frac{n-2}{n}\sum_1^n |z_j|^2, \tag{3.2}$$

with the equality sign if and only if all z_j *are real.*

Proof. By (1.9) we have

$$\sum_1^{n-1} w_k^2 = \frac{n-2}{n}\sum_1^n z_j^2,$$

and therefore

$$\sum_1^{n-1} w_k^2 = \frac{n-2}{n}\left|\sum_1^n z_j^2\right| \leq \frac{n-2}{n}\sum_1^n |z_j|^2. \tag{3.3}$$

This proves (3.2).

If we have the equality sign in (3.2), then it surely holds also in (3.3), and this implies that

$$\left|\sum_1^n z_j^2\right| = \sum_1^n |z_j|^2.$$

This last equation implies that all n complex numbers z_j^2 have the same argument. But this means that there is an angle θ so that $z_j^2 = |z_j|^2 e^{i\theta}$. This, however, implies that

$$z_j = \pm |z_j| e^{i\theta/2} \quad \text{for all } j,$$

and it follows that all points z_j are on a line L through O.

By the Gauss-Lucas theorem the convex hull of the z_j must contain the convex hull of the w_k. As the w_k are all real by assumption, it follows that we must have $\theta = 0$, hence all z_j are real. This completes our proof of Theorem 3.

4. Verifying Conjecture 1 for binomial polynomials. We say that our polynomial is *binomial* provided that it is of the form $P_n(z) = z^n + a_k z^{n-k}$, where we may as well assume that $a_k = 1$; hence

$$P_n(z) = z^n + z^{n-k} = z^{n-k}(z^k + 1), \quad (2 \leq k \leq n), \tag{4.1}$$

and

$$P_n'(z) = nz^{n-1} + (n-k)z^{n-k-1} = nz^{n-k-1}\left(z^k + \frac{n-k}{n}\right). \tag{4.2}$$

Denoting, as before, their zeros by z_j and w_k, we now find that

$$\sum_1^{n-1} |w_2|^2 = k\left(\frac{n-k}{n}\right)^{2/k} \quad \text{and} \quad \sum_1^n |z_j|^2 = k,$$

so that (1.11) amounts to the inequality

$$k\left(\frac{n-k}{n}\right)^{2/k} \leqslant \frac{n-2}{n}k.$$

This being evident if $k = 2$ or if $k = n$, there remains to prove

LEMMA 1. *We have the inequality*

$$(n-2)^k > n^{k-2}(n-k)^2 \quad \text{if} \quad 2 < k < n. \tag{4.3}$$

This we derive from the more general

LEMMA 2. *If the reals $x_1, \dots, x_k$, not all equal to each other, have the arithmetic mean*

$$a = \frac{1}{k}\sum_1^k x_j, \tag{4.4}$$

then

$$(x-a)^k > \prod_1^k (x - x_j) \quad \text{if} \quad x > x_j \ (j = 1, \dots, k). \tag{4.5}$$

Proof of Lemma 2. Taking logarithms, we see that (4.5) is equivalent to

$$\log(x-a) > \frac{1}{k}\sum_1^k \log(x - x_j) \quad \text{if} \quad x > \max x_j. \tag{4.6}$$

From (4.4) we have

$$x - a = \frac{1}{k}\sum_1^k (x - x_j)$$

and (4.6) amounts to

$$\log\left(\frac{1}{k}\sum_1^k (x - x_j)\right) > \frac{1}{k}\sum_1^k \log(x - x_j), \tag{4.7}$$

which follows from the *strict concavity* of $\log x$ in $0 < x < \infty$. Indeed, for any *strictly concave* function $f(x)$ in $(0, \infty)$, and for positive quantities $p_j = x - x_j$, not all equal to each other, we have the well-known inequality

$$f\left(\frac{1}{k}\sum_1^k p_j\right) > \frac{1}{k}\sum_1^k f(p_j).$$

For $f(x) = \log x$ this amounts to (4.7).

We now specialize Lemma 2 by choosing

$$x_1 = x_2 = \cdots = x_{k-2} = 0, \qquad x_{k-1} = x_k = k, \quad \text{and} \quad x = n.$$

For the mean value (4.4) we find

$$a = \frac{1}{k}\sum_1^k x_j = \frac{1}{k} \cdot 2k = 2,$$

while

$$\prod_{1}^{k-2}(x-x_j)=n^{k-2}, \qquad (x-x_{k-1})(x-x_k)=(n-k)^2.$$

Now (4.5) goes over into the desired inequality (4.3).

References

1. F. J. van den Berg, Nogmaals over afgeleide Wortelpunten, Nieuw Archief voor Wiskunde, 15 (1888) 100–164.

2. Morris Marden, Geometry of Polynomials, Mathematical Surveys, no. 3, Amer. Math. Society, 1966.

3. I. J. Schoenberg, Mathematical Time Exposures, Mathematical Association of America, Washington, DC, 1982.

THE INTERVAL OF CONVERGENCE AND LIMITING FUNCTIONS OF A HYPERPOWER SEQUENCE

J. M. DE VILLIERS AND P. N. ROBINSON

Department of Applied Mathematics, University of Stellenbosch, 7600 Stellenbosch, South Africa

1. Introduction. In this article we study the convergence of the infinitely iterated exponential

$$x=h(z)=z^{z^{z^{\cdot^{\cdot^{\cdot}}}}} \tag{1.1}$$

for real positive z. We prove that $h(z)$ exists for $z\in[e^{-e},e^{1/e}]$ by showing that $x=h(z)$ is the unique solution of the *first auxiliary equation*

$$f(z,x)\equiv x-z^x=0 \tag{1.2}$$

for $(z,x)\in(1,e^{1/e}]\times[0,e]$, and also the unique solution of the *second auxiliary equation*

$$g(z,x)\equiv x-z^{z^x}=0 \tag{1.3}$$

for $(z,x)\in[e^{-e},1)\times[0,\infty)$. In addition we prove that the sequence of hyperpowers implied by the right hand side of (1.1) diverges for $z\notin[e^{-e},e^{1/e}]$. For $(z,x)\in(0,e^{-e})\times[0,\infty)$ it turns out that the second auxiliary equation $g(z,x)=0$ has exactly three solutions $s_1(z)$, $s(z)$ and $s_2(z)$ satisfying

$$0<s_1(z)<s(z)<s_2(z)<1,$$

where $x=s(z)$ also solves $f(z,x)=0$, and we show here that the subsequences of odd and even hyperpowers converge to $s_1(z)$ and $s_2(z)$, respectively, hence the divergence of the original hyperpower sequence. For $z\in(e^{1/e},\infty)$ we prove that divergence is a direct consequence of the non-existence of solutions of the first auxiliary equation $f(z,x)=0$.

The same interval of convergence was established in an article by Knoebel [2], where the implicit function theorem was also employed, but where a proof by contradiction was used to prove the divergence of the hyperpower sequence for $z\in(0,e^{-e})$. Our more direct approach leads to estimates that are sharper than some of those obtained there, and yields in addition some interesting properties of the limiting functions of the hyperpower sequence (1.1).

J. M. de Villiers: I obtained a Ph.D. degree at Cambridge University in 1974 and was then appointed to the Department of Applied Mathematics at the University of Stellenbosch (about 50 kilometers from Cape Town), where I now hold the position of associate professor. The courses I teach, as well as my research interests, are mainly in the fields of applied functional analysis, differential equations and numerical analysis. The second author, P. N. Robinson, was a graduate student under my supervision at this department. For the last eight years I have been, in addition to my mathematical teaching duties, full-time conductor of the Stellenbosch University Choir. At the end of 1984 I laid down my post as choral conductor in order to devote more time to mathematical teaching and research.

For an account of the history and background of this problem, as well as an extensive bibliography, we refer the reader to Knoebel's article [2].

2. The main result. Adopting the convention

$$0^0 = 1, \tag{2.1}$$

we define for real nonnegative z the sequence $\{z_n\}$ by

$$z_1 = z, \quad z_{n+1} = z^{z_n} \quad \text{for } n = 1, 2, \ldots,$$

and use the notation

$$^n z \equiv z_n, \quad n = 1, 2, \ldots$$

for the resulting hyperpowers of z. Whenever the sequence $\{^n z\}$ converges, we write

$$h(z) \equiv \lim_{n \to \infty} {}^n z.$$

The constants e, e^{-1}, e^{-e} and $e^{1/e}$ will appear frequently in our subsequent results, so that it is convenient to know their approximate numerical values:

$$e \doteq 2.718281828 \qquad e^{-1} \doteq 0.3678794412$$
$$e^{-e} \doteq 0.06598803585 \qquad e^{1/e} \doteq 1.444667861.$$

Our main result is then the following:

THEOREM. (a) *The sequence* $\{^n z\}$ *converges for* $z \in [e^{-e}, e^{1/e}]$, *where the limiting function* $h(z) \equiv \lim_{n \to \infty} {}^n z$ *enjoys the following properties:*

(i) $h(z)$ *is real analytic and strictly increasing for* $z \in (e^{-e}, e^{1/e})$, *with*

$$\lim_{z \to e^{1/e}} h'(z) = \infty.$$

(ii) $h(z)$ *is continuous for* $z \in [e^{-e}, e^{1/e}]$, *with*

$$h(e^{-e}) = e^{-1}, \quad h(e^{1/e}) = e.$$

(b) *The sequence* $\{^n z\}$ *diverges for* $z \in (e^{1/e}, \infty)$, *with*

$$\lim_{n \to \infty} {}^n z = \infty.$$

(c) *The sequence* $\{^n z\}$ *diverges for* $z \in [0, e^{-e})$, *with the odd and even subsequences converging to different values*:

$$\lim_{n \to \infty} {}^{(2n+1)}z \equiv h_1(z), \quad \lim_{n \to \infty} {}^{(2n+2)}z \equiv h_2(z),$$

where

$$0 < h_1(z) < e^{-1} < h_2(z) < 1 \quad \text{for} \quad z \in (0, e^{-e}),$$

and where the limiting functions $h_1(z)$ *and* $h_2(z)$ *have the following properties*:

(i) $h_1(z)$ *and* $h_2(z)$ *are real analytic for* $z \in (0, e^{-e})$, *with* $h_1(z)$ *strictly increasing and* $h_2(z)$ *strictly decreasing for* $z \in (0, e^{-e})$.

(ii) $h_1(z)$ *and* $h_2(z)$ *are continuous for* $z \in [0, e^{-e}]$, *with*

$$h_1(0) = 0, \quad h_2(0) = 1,$$

and

$$h_1(e^{-e}) = h_2(e^{-e}) = h(e^{-e}) = e^{-1}.$$

The properties of the limiting functions $h(z)$, $h_1(z)$ and $h_2(z)$, as stated above, are illustrated by the graphs drawn in Fig. 1. In particular, the bifurcation occurring at the point $(z, x) = (e^{-e}, e^{-1})$ is clearly shown.

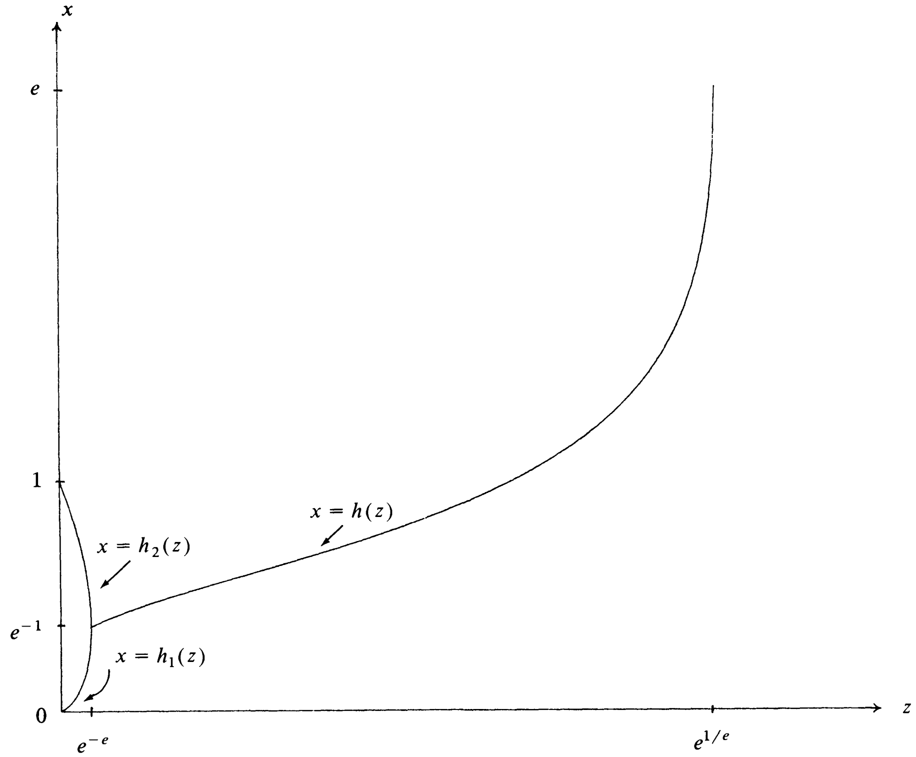


FIG. 1. The limiting functions $x = h(z)$, $x = h_1(z)$ and $x = h_2(z)$. The bifurcation at the point $(z, x) = (e^{-e}, e^{-1})$ is clearly shown.

3. Solutions of the auxiliary equations. The proof of the theorem depends on the following two lemmas which are concerned with the existence, uniqueness and other relevant properties of solutions of the two auxiliary equations $f(z, x) = 0$ and $g(z, x) = 0$, as given by (1.2) and (1.3).

LEMMA 3.1. (a) *The first auxiliary equation* $f(z, x) = 0$ *has for* $(z, x) \in (0, e^{1/e}] \times [0, e]$ *a unique solution* $x = s(z)$ *with the following properties*:

(i) $s(z)$ *is real analytic for* $z \in (0, e^{1/e})$.

(ii) $s'(z) > 0$ *for* $z \in (0, e^{1/e})$.

(iii) $s(z)$ *is continuous for* $z \in (0, e^{1/e}]$, *with*

$$s(e^{-e}) = e^{-1}, \quad s(e^{1/e}) = e.$$

(iv) $\lim_{z \to e^{1/e}} s'(z) = \infty$.

(b) *The equation* $f(z, x) = 0$ *has no solution for* $(z, x) \in (e^{1/e}, \infty) \times [0, \infty)$.

LEMMA 3.2. (a) *The second auxiliary equation* $g(z, x) = 0$ *has for* $(z, x) \in [e^{-e}, 1) \times [0, \infty)$ *a unique solution* $x = s(z)$, *where* $s(z)$ *is the same function as in Lemma* 3.1(*a*).

(b) *The equation* $g(z, x) = 0$ *has for* $(z, x) \in (0, e^{-e}) \times [0, \infty)$ *exactly three solutions* $x = s_1(z)$, $x = s(z)$ *and* $x = s_2(z)$, *where* $s(z)$ *is the same function as in Lemma* 3.1(*a*), *and where the functions* $s_1(z)$ *and* $s_2(z)$ *enjoy the following properties*:

(3.1)

(i) $0 < s_1(z) < s(z) < e^{-1} < s_2(z) < 1$ *for* $z \in (0, e^{-e})$.

(ii) $s_1(z)$ *and* $s_2(z)$ *are real analytic for* $z \in (0, e^{-e})$.

(iii) $s_1'(z) > 0$ *for* $z \in (0, e^{-e})$.

(iv) $s_2'(z) < 0$ *for* $z \in (0, e^{-e})$.

(v) $s_1(z)$ *and* $s_2(z)$ *are continuous for* $z \in (0, e^{1/e}]$, *with*

$$s_1(e^{-e}) = s_2(e^{-e}) = s(e^{-e}) = e^{-1}.$$

(3.2)

(vi) $\lim_{z \to 0} s_1(z) = 0, \lim_{z \to 0} s_2(z) = 1.$

Before proceeding to the proofs of the lemmas, we state the following three inequalities which are necessary for our further work.

INEQUALITIES. For real positive a, b and c it can easily be shown that

(3.3a) if $a < b$, then $a^c < b^c$;
(3.3b) if $a < b$, and $c < 1$, then $c^a > c^b$;
(3.3c) if $a < b$, and $c > 1$, then $c^a < c^b$.

Proof of Lemma 3.1(a). To solve the equation $f(z,x) = 0$ for x, we fix $z \in (0, e^{1/e})$, write $F(x)$ for $f(z,x)$ so that now

$$F(x) \equiv x - z^x, \tag{3.4}$$

and examine the existence of zeroes of the function $F(x)$ on the interval $[0, e]$. Then, as illustrated by the special case drawn in Fig. 2, we have $F(0) < 0$, $F(e) > 0$, and

$$F''(x) = -z^x(\ln z)^2 < 0 \quad \text{for} \quad x \in [0, e].$$

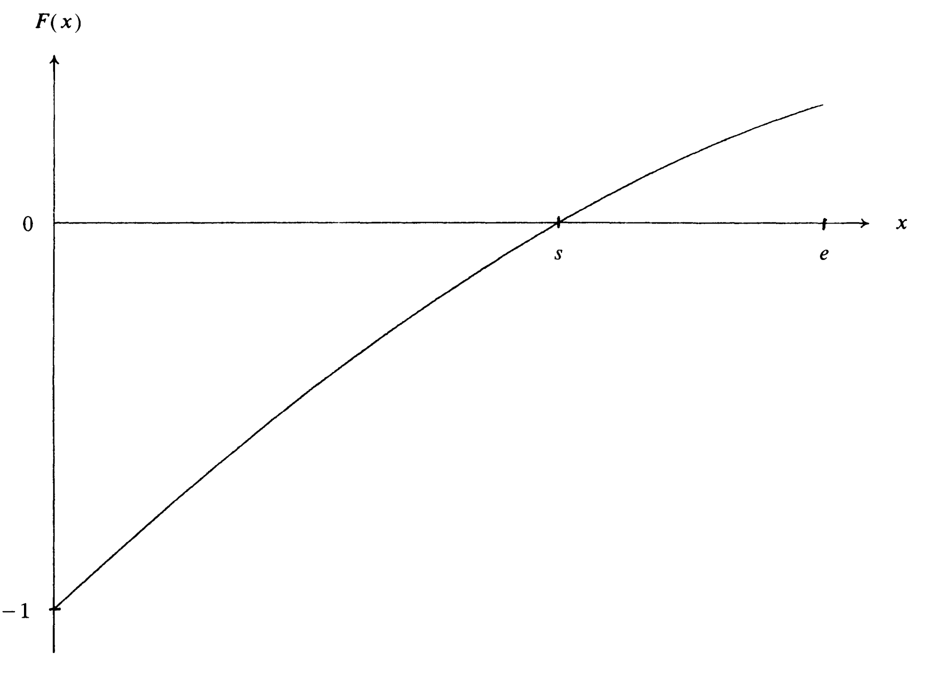


FIG. 2. The function $F(x) \equiv f(1.380, x)$. Here $s \doteq 1.766337$.

It follows that $F(x)$ has a unique zero $x = s$ in the interval $(0, e)$, with $F'(s) > 0$. Hence the equation $f(z,x) = 0$ has for $(z,x) \in (0, e^{1/e}) \times [0, e]$ the unique solution $x = s(z)$, where we have now written $s(z)$ for s. It is also clear that

$$0 < s(z) < e \quad \text{for} \quad z \in (0, e^{1/e}), \tag{3.5}$$

and

$$\left\{ \frac{\partial}{\partial x} f(z,x) \right\} \Bigg|_{x=s(z)} > 0 \quad \text{for} \quad z \in (0, e^{1/e}). \tag{3.6}$$

Next, we investigate the behaviour of the function $s(z)$ in the interval $(0, e^{1/e})$. Observing that $s(z)$ satisfies the relation

$$f(z,s(z)) \equiv s(z) - z^{s(z)} = 0 \quad \text{for} \quad z \in (0, e^{1/e}), \tag{3.7}$$

and noting also (3.6), we employ the implicit function theorem ([**3**], p. 56) to deduce that $s(z)$ is continuously differentiable for $z \in (0, e^{1/e})$. Implicit differentiation of the relation (3.7) now yields the formula

$$s'(z) = z^{-1}s^2(z)\{1 - \ln s(z)\}^{-1} \quad \text{for} \quad z \in (0, e^{1/e}), \tag{3.8}$$

where, by virtue of (3.5),

$$1 - \ln s(z) > 0 \quad \text{for} \quad z \in (0, e^{1/e}),$$

from which we immediately obtain the strict monotone property (ii) of $s(z)$.

Noting that the right-hand side of (3.8) is real analytic in s and z, we now regard $s(z)$ as a solution of the ordinary differential equation (3.8), and appeal to standard theory ([**1**], p. 103) to deduce the desired real analyticity of $s(z)$ for $z \in (0, e^{1/e})$.

Observing that a simple calculation verifies the fact that $s(e^{-e}) = e^{-1}$, we proceed to set $z = e^{1/e}$ in the definition (3.4) of $F(x)$. But then $F(0) < 0, F(e) = 0$ and

$$F'(x) = 1 - e^{-1}e^{e^{-1}x} > 0$$

for $x \in [0, e)$, with $F'(e) = 0$, from which we deduce that $x = e$ solves the equation $f(e^{1/e}, x) = 0$ uniquely on $[0, e]$. Thus $s(e^{1/e}) = e$, so that to prove the continuity of $s(z)$ for $z \in (0, e^{1/e}]$, it will suffice to show that the limit

$$\lim_{z \to e^{1/e}} s(z) \equiv l$$

exists, with $l = e$. But the properties (i) and (ii) of $s(z)$, together with (3.5), imply that, for $z \in (0, e^{1/e})$, the function $s(z)$ is continuous, strictly increasing and bounded above by e. Hence the limit l does indeed exist, with $l \leqslant e$. Moreover, by letting $z \to e^{1/e}$ in the relation (3.7), we find that l satisfies $f(e^{1/e}, l) = 0$, yielding the desired $l = s(e^{1/e}) = e$.

Finally, the limiting property (iv) of $s'(z)$ follows if we insert the fact that $s(z) \to e$ as $z \to e^{1/e}$ into the formula (3.8) for $s'(z)$, keeping in mind also that $s(z)$ is strictly increasing for $z \in (0, e^{1/e})$.

(b) We prove the non-existence of solutions of the equation $f(z, x) = 0$ for $(z, x) \in (e^{1/e}, \infty) \times [0, \infty)$ by showing that, for a fixed $z \in (e^{1/e}, \infty)$, the function $F(x)$ of (3.4) is strictly negative on $[0, \infty)$. First, some elementary calculus reveals that $F(x)$ attains its maximum value over $[0, \infty)$ at the point $x = x_m$, where

$$x_m = (\ln z)^{-1}\ln\{(\ln z)^{-1}\}.$$

This yields the maximum

$$F(x_m) = (\ln z)^{-1}\left[\ln\{(\ln z)^{-1}\} - 1\right] < 0$$

for $z \in (e^{1/e}, \infty)$, as can be shown by some routine analysis. Q.E.D.

Proof of Lemma 3.2(a). The solution $x = s(z)$ of the first auxiliary equation $f(z, x) = 0$ clearly also solves the second auxiliary equation $g(z, x) = 0$ for $(z, x) \in [e^{-e}, 1) \times [0, \infty)$. To investigate the uniqueness of this solution, we fix $z \in (e^{-e}, 1)$, write $G(x)$ for $g(z, x)$ so that now

$$G(x) \equiv x - z^{z^x}, \tag{3.9}$$

and proceed to examine the properties of the function $G(x)$ on $[0, \infty)$. Since clearly $G(s) = 0$, the uniqueness of the solution $x = s(z)$ of $g(z, x) = 0$ for $(z, x) \in (e^{-e}, 1) \times [0, \infty)$ will follow if we can prove that

(3.10) $$G'(x) > 0 \text{ on } [0, \infty) \quad \text{for} \quad z \in (e^{-e}, 1).$$

To find the minimum value of $G'(x)$ on $[0, \infty)$ we differentiate (3.9) twice, obtaining

(3.11) $$G'(x) = 1 - z^{z^x} z^x (\ln z)^2$$

and

(3.12) $$G''(x) = -z^{z^x} z^x (\ln z)^3 (\ln z^{z^x} + 1),$$

and then equate the right-hand side of (3.12) to zero. This procedure leads to the conclusion that $G''(x)$ has a zero $x = s_0$ in $(0, \infty)$ only if there exists a point $s_0 \in (0, \infty)$ satisfying

(3.13a) $$z^{z^{s_0}} = e^{-1},$$

in which case s_0 is uniquely determined by

(3.13b) $$s_0 = (\ln z)^{-1} \ln\{-(\ln z)^{-1}\}.$$

Now restrict z to the interval $[e^{-1}, 1)$. From (3.12) and (3.13a) it is then clear that $G''(x) \geqslant 0$ in $[0, \infty)$, yielding the bound

(3.14) $$\min_{x \in [0, \infty)} G'(x) = G'(0) = 1 - z(\ln z)^2 \geqslant 1 - e^{-1} \quad \text{for} \quad z \in [e^{-1}, 1),$$

as can easily be verified.

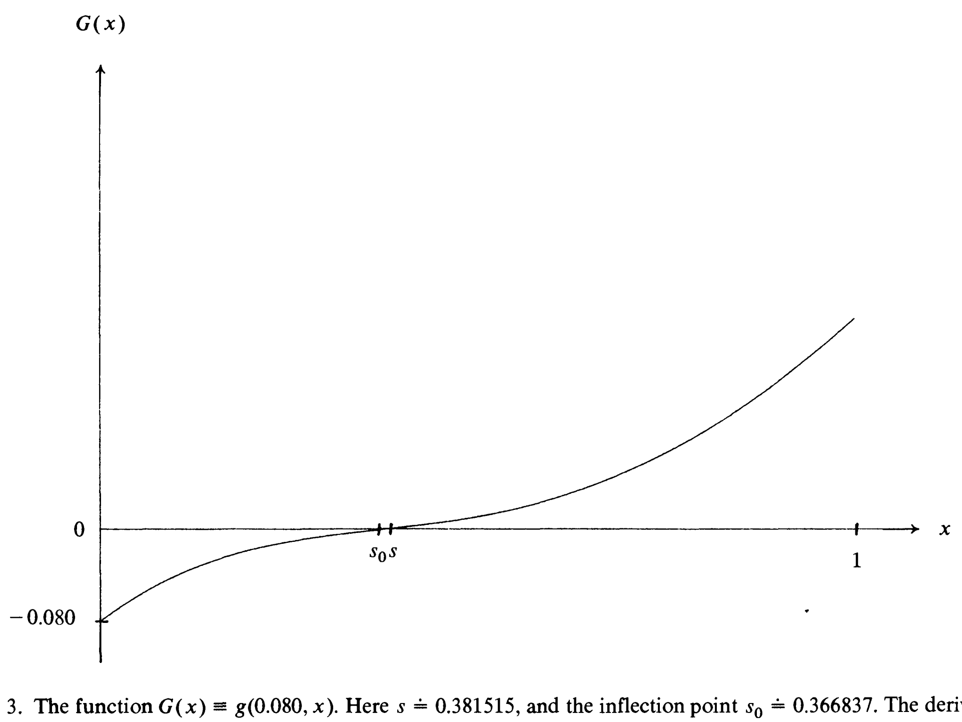


FIG. 3. The function $G(x) \equiv g(0.080, x)$. Here $s \doteq 0.381515$, and the inflection point $s_0 \doteq 0.366837$. The derivative $G'(x)$ is clearly strictly positive in the interval shown.

Next we consider the case $z \in (e^{-e}, e^{-1})$. Then the turning point s_0 as given by (3.13a, b) clearly belongs to $(0, \infty)$. Also $G''(x) < 0$ for $x \in (0, s_0)$ and $G''(x) > 0$ for $x \in (s_0, \infty)$, so that (see also Fig. 3)

(3.15) $$\min_{x \in [0, \infty)} G'(x) = G'(s_0) = 1 + e^{-1} \ln z > 0 \quad \text{for} \quad z \in (e^{-e}, e^{-1}).$$

The desired positivity (3.10) of $G'(x)$ now follows from (3.14) and (3.15).

Finally we set $z = e^{-e}$ in the definition (3.9) of $G(x)$ to prove uniqueness for this special case. As illustrated in Fig. 4, a routine check now reveals that the function $G(x)$ has for $x \in [0, \infty)$ a unique turning point and a unique inflection point coinciding at the solution $x = s(e^{-e}) = e^{-1}$ of the equation $g(e^{-e}, x) = 0$. These findings suffice to imply uniqueness for the case $z = e^{-e}$.

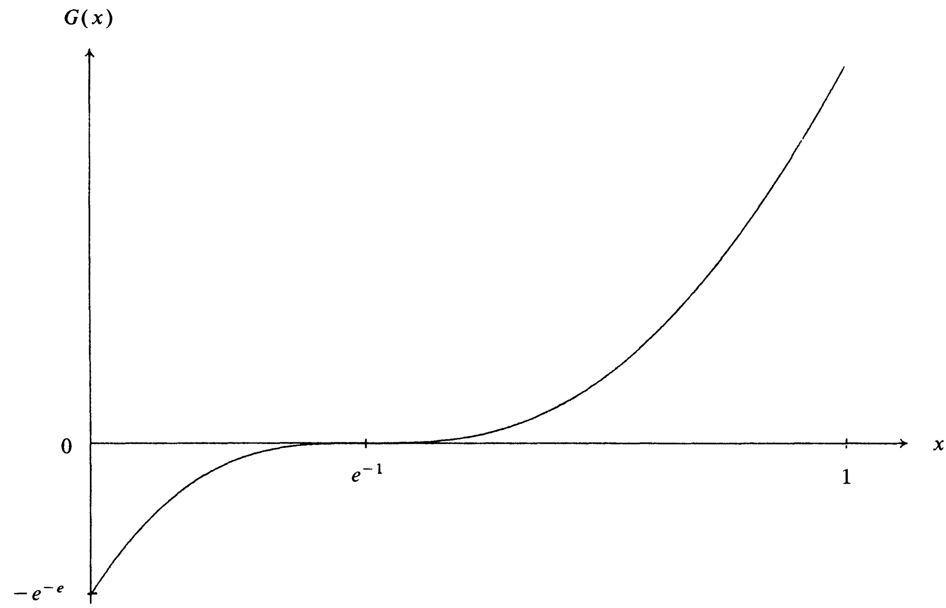


FIG. 4. The function $G(x) \equiv g(e^{-e}, x)$.

(b) The function $x = s(z)$ of Lemma 3.1(a) clearly also solves the equation $g(z, x) = 0$ for $(z, x) \in (0, e^{-e}) \times [0, \infty)$. Moreover, since $s(z)$ is strictly increasing, with $s(e^{-e}) = e^{-1}$, we have the bounds

$$0 < s(z) < e^{-1} \quad \text{for} \quad z \in (0, e^{-e}). \tag{3.16}$$

To examine the existence of additional solutions of $g(z, x) = 0$, we fix $z \in (0, e^{-e})$, define the function $G(x)$ by (3.9), and study the behaviour of its zeroes on $[0, \infty)$.

Writing s, s_1 and s_2 for $s(z)$, $s_1(z)$ and $s_2(z)$, respectively, whenever convenient, we first restrict x to the interval $[0, s]$. Then, as illustrated in the example drawn in Fig. 5, we clearly have $G(0) < 0$ and $G(s) = 0$, with

$$G'(s) = 1 - z^{z^s} z^s (\ln z)^2 = 1 - (\ln s)^2 < 0 \tag{3.17}$$

by virtue of (3.11), (3.7) and (3.16). In addition, we have from (3.12) that

$$G''(x) = -z^{z^x} z^x (\ln z)^3 (z^x \ln z + 1) < 0 \quad \text{for} \quad x \in [0, s], \tag{3.18}$$

where we have used the inequalities (3.3b) and (3.16) to show that

$$z^x \ln z + 1 \leqslant z^s \ln z + 1 = \ln s + 1 < 0 \quad \text{for} \quad x \in [0, s].$$

It follows that $G(x)$ has a unique zero $x = s_1$ in the interval $(0, s)$, with

$$G'(s_1) > 0. \tag{3.19}$$

Next we consider the function $G(x)$ on the interval $[s, \infty)$. Then (see also Fig. 5) we clearly have $G(s) = 0$, $G'(s) < 0$ and $G''(s) < 0$ from (3.17) and (3.18). Moreover, using the same arguments as after (3.12) above, we deduce that $G(x)$ has exactly one inflection point $x = s_0$ in the interval (s, ∞) with s_0 given by (3.13a, b). Since also $G(1) > 0$, it follows that $G(x)$ has a

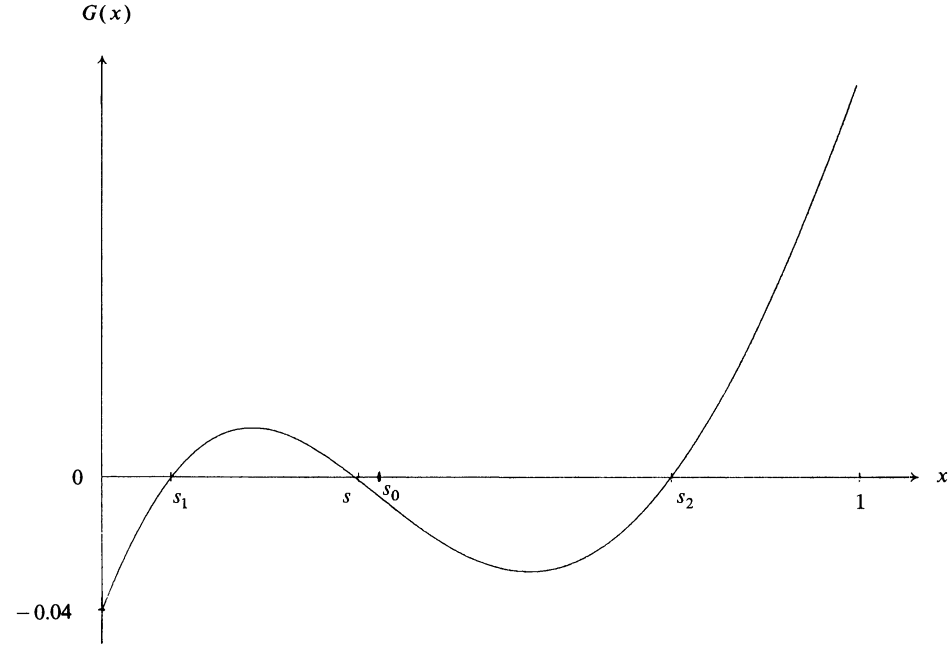


FIG. 5. The function $G(x) \equiv g(0.04, x)$. Here $s_1 \doteq 0.089601$, $s \doteq 0.337471$, $s_2 \doteq 0.749451$, and the inflection point $s_0 \doteq 0.363180$.

unique zero $x = s_2$ in the interval (s, ∞), with $s_0 < s_2 < 1$, and where

$$G'(s_2) > 0. \tag{3.20}$$

We have now established that the equation $g(z,x) = 0$ has for $(z,x) \in (0, e^{-e}) \times [0, \infty)$ exactly three solutions $x = s_1(z), x = s(z)$ and $x = s_2(z)$, with

$$0 < s_1(z) < s(z) < s_0(z) < s_2(z) < 1 \quad \text{for} \quad z \in (0, e^{-e}), \tag{3.21}$$

and where $s_0 = s_0(z)$ is given by (3.13b). In addition, we can write

$$s_2(z) - e^{-1} = z^{z^{s_2(z)}} - z^{z^{s_0(z)}} = \int_{s_0(z)}^{s_2(z)} z^{z^t} z^t (\ln z)^2 \, dt > 0,$$

which, together with (3.21), verify the inequalities (3.1).

The real analyticity of $s_1(z)$ and $s_2(z)$ is proved by means of an argument similar to the one used in the proof of Lemma 3.1(a). The functions $s_1(z)$ and $s_2(z)$ clearly satisfy the relation

$$g(z, s_i(z)) \equiv s_i(z) - z^{z^{s_i(z)}} = 0 \quad \text{for} \quad i = 1,2 \quad \text{and} \quad z \in (0, e^{-e}), \tag{3.22}$$

and, by virtue of (3.19) and (3.20), it is also clear that

$$\left\{ \frac{\partial}{\partial x} g(z,x) \right\} \bigg|_{x = s_i(z)} > 0 \quad \text{for} \quad i = 1,2 \quad \text{and} \quad z \in (0, e^{-e}).$$

We may therefore apply the implicit function theorem as before to deduce that the functions $s_1(z)$ and $s_2(z)$ are continuously differentiable for $z \in (0, e^{-e})$. Implicit differentiation of the relation (3.22) then yields the formula

$$s_i'(z) = s_i(z) z^{s_i(z)-1} \{1 + s_i(z) \ln z\} \{1 - s_i(z)(\ln z) \ln s_i(z)\}^{-1} \quad \text{for} \quad i = 1,2, \tag{3.23}$$

with

$$1 - s_i(z)(\ln z) \ln s_i(z) = G'(s_i) > 0 \quad \text{for} \quad i = 1,2, \tag{3.24}$$

from (3.19) and (3.20). Now regard $s_i(z)$ as a solution of the analytic ordinary differential equation (3.23), and appeal to standard theory ([1], p. 103) as before to deduce that $s_1(z)$ and $s_2(z)$ are real analytic for $z \in (0, e^{-e})$.

The strict monotone property (iii) of $s_1(z)$ now follows from (3.23) and (3.24), together with the fact that

$$1 + s_1(z)\ln z > 1 + \{\ln s_1(z)\}^{-1} > 0,$$

where we have used (3.24) and the inequality $s_1(z) < e^{-1}$.

Similarly, property (iv) of s_2 is proved by noting that, since $s_2(z) > e^{-1}$, we have

$$1 + s_2(z)\ln z < 1 + e^{-1}\ln z < 0 \quad \text{for} \quad z \in (0, e^{-e}),$$

and then combining this fact with (3.23) and (3.24).

The continuity of $s_1(z)$ and $s_2(z)$ on the half-open interval $(0, e^{-e}]$ is now proved by means of a similar technique to the one used to show the analogous property (a) (iii) of $s(z)$ in the proof of Lemma 3.1(a). First, the properties (i)–(iv) of $s_1(z)$ and $s_2(z)$ can be shown to imply that the limits

$$\lim_{z \to e^{-e}} s_1(z) \equiv l_1, \ \lim_{z \to e^{-e}} s_2(z) \equiv l_2$$

exist, with $l_1 \leqslant e^{-1}$ and $l_2 \geqslant e^{-1}$. If we now let $z \to e^{-e}$ in the relation (3.22), we find that l_1 and l_2 satisfy $g(e^{-e}, l_1) = 0$ and $g(e^{-e}, l_2) = 0$. But we have noted in Lemma 3.2(a) that $x = s(e^{-e}) = e^{-1}$ solves $g(e^{-e}, x) = 0$ uniquely on $[0, \infty)$. Thus $l_1 = l_2 = e^{-1}$, verifying property (v) of $s_1(z)$ and $s_2(z)$.

The limiting property (vi) of $s_1(z)$ for $z \to 0$ will clearly follow from (3.21) if we can show that

$$\lim_{z \to 0} s_0(z) = 0. \tag{3.25}$$

We therefore investigate this limit by applying L'Hospital's rule to the right hand side of (3.13b), obtaining

$$\lim_{z \to 0} \frac{\ln\{-(\ln z)^{-1}\}}{\ln z} = \lim_{z \to 0} (-\ln z)^{-1} = 0,$$

which yields the desired (3.25).

The analogous limit of $s_2(z)$ as $z \to 0$ is obtained by first noting that $e^{-1} < s_2(z) < 1$ for $z \in (0, e^{-e})$ from (3.1), and then writing

$$z^z - s_2(z) = \int_{s_2(z)}^1 z^{z^t} z^t (\ln z)^2 \, dt < \int_{e^{-1}}^1 z^t (\ln z)^2 \, dt \leqslant z^{e^{-1}} (\ln z)^2 (1 - e^{-1})$$

for $z \in (0, e^{-e})$, so that

$$1 > s_2(z) > z^z - (1 - e^{-1}) z^{e^{-1}} (\ln z)^2 \to 1 \quad \text{as } z \to 0. \text{ Q.E.D.}$$

4. Proof of theorem. Our main result can now be established by means of a procedure whereby the limiting functions $h(z)$, $h_1(z)$ and $h_2(z)$ of the hyperpower sequence $\{{}^n z\}$ are matched to the corresponding solutions $s(z)$, $s_1(z)$ and $s_2(z)$, respectively, of the auxiliary equations $f(z, x) = 0$ and $g(z, x) = 0$.

Proof of theorem. (a) Suppose first that $z \in (1, e^{1/e}]$. Using (3.3c) we can then show that $\{{}^n z\}$ is a strictly increasing sequence. Moreover we can prove by mathematical induction that the sequence $\{{}^n z\}$ is bounded above:

$${}^n z \leqslant e \quad \text{for} \quad n = 1, 2, 3, \ldots \quad \text{and} \quad z \in (1, e^{1/e}].$$

Therefore $\{{}^n z\}$ has a limit:

$$\lim_{n \to \infty} {}^n z \equiv h(z),$$

with

$$0 < h(z) \leqslant e \quad \text{for} \quad z \in (1, e^{1/e}]. \tag{4.1}$$

Now use the fact that z^x is continuous in x to obtain

$$h(z) = \lim_{n\to\infty} {}^{(n+1)}z = \lim_{n\to\infty} z^{{}^n z} = z^{\lim_{n\to\infty} {}^n z} = z^{h(z)},$$

from which it follows that $f(z, h(z)) = 0$ for $z \in (1, e^{1/e}]$. Noting also (4.1), we then deduce from Lemma 3.1(a) that

$$h(z) = s(z) \quad \text{for} \quad z \in (1, e^{1/e}],$$

thereby establishing the result (a) of the theorem for these values of z.

Observing that the case $z = 1$ is trivial, we suppose next that $z \in (0,1)$. Using the inequalities (3.3a, b) we can now show that the odd subsequence $\{{}^{(2n+1)}z\}$ is strictly increasing, and the even subsequence $\{{}^{(2n+2)}z\}$ strictly decreasing. But $\{{}^{(2n+1)}z\}$ is clearly bounded above by one and $\{{}^{(2n+2)}z\}$ bounded below by zero, so that both subsequences have limits:

$$\lim_{n\to\infty} {}^{(2n+1)}z \equiv h_1(z), \quad \lim_{n\to\infty} {}^{(2n+2)}z \equiv h_2(z) \quad \text{for} \quad z \in (0,1), \tag{4.2}$$

$$0 < h_1(z) \leqslant 1, \quad 0 \leqslant h_2(z) < 1 \quad \text{for} \quad z \in (0,1). \tag{4.3}$$

Now exploit the continuity of z^{z^x} as a function of x to obtain

$$h_1(z) = \lim_{n\to\infty} {}^{(2n+3)}z = \lim_{n\to\infty} z^{z^{(2n+1)_z}} = z^{z^{\lim_{n\to\infty} {}^{(2n+1)}z}} = z^{z^{h_1(z)}} \quad \text{for} \quad z \in (0,1),$$

and similarly that

$$h_2(z) = z^{z^{h_2(z)}} \quad \text{for} \quad z \in (0,1).$$

Hence $h_1(z)$ and $h_2(z)$ satisfy the relations

$$g(z, h_1(z)) = 0, \quad g(z, h_2(z)) = 0 \quad \text{for} \quad z \in (0,1). \tag{4.4}$$

If we restrict z to the interval $[e^{-e}, 1)$, it then follows from Lemma 3.2(a) and (4.3) that

$$h_1(z) = s(z) = h_2(z) \quad \text{for} \quad z \in [e^{-e}, 1),$$

which, together with the properties (i)–(iv) of $s(z)$ given in Lemma 3.1(a), conclude the proof of part (a) of the theorem.

(b) Let $z \in (e^{1/e}, \infty)$ and suppose $\{{}^n z\}$ has a finite limit:

$$\lim_{n\to\infty} {}^n z \equiv l_0(z) < \infty.$$

But then, arguing by the continuity of exponentiation as before, it would follow that

$$l_0(z) = z^{l_0(z)} \quad \text{for} \quad z \in (e^{1/e}, \infty),$$

which contradicts Lemma 3.1(b). Since $\{{}^n z\}$ is strictly increasing by virtue of (3.3c), it follows that $\lim_{n\to\infty} {}^n z = \infty$ for $z \in (e^{1/e}, \infty)$.

(c) For $z \in (0, e^{-e})$ we use the inequalities

$$0 < s_1(z) < e^{-1} < s_2(z) < 1$$

from (3.1), together with (3.22), to obtain

$$s_1(z) - z = z^{z^{s_1(z)}} - z = \int_0^{s_1(z)} z^{z^t} z^t (\ln z)^2 \, dt > 0,$$

and, similarly,

$$s_2(z) - z^z = z^{z^{s_2(z)}} - z^z = \int_1^{s_2(z)} z^{z^t} z^t (\ln z)^2 \, dt < 0,$$

so that the statement

$$(4.5) \qquad {}^{(2n+1)}z < s_1(z) < e^{-1} < s_2(z) < {}^{(2n+2)}z \quad \text{for} \quad n = 0,1,2,\dots$$

is true for $n = 0$. The validity of (4.5) for all n now follows by mathematical induction, the inductive steps from $n = k$ to $n = k + 1$ being

$${}^{(2k+3)}z = z^{{}^{(2k+1)}z} < z^{z^{s_1(z)}} = s_1(z)$$

and

$${}^{(2k+4)}z = z^{{}^{(2k+2)}z} > z^{z^{s_2(z)}} = s_2(z),$$

where we have used (3.3b) repeatedly.

Next, combine (4.2) and (4.5) to obtain the estimates

$$(4.6) \qquad 0 < h_1(z) \leqslant s_1(z) < e^{-1} < s_2(z) \leqslant h_2(z) < 1 \quad \text{for} \quad z \in (0, e^{-e}).$$

Moreover, $h_1(z)$ and $h_2(z)$ satisfy the relations (4.4), so that we may use Lemma 3.2(b) to conclude from (4.6) that

$$(4.7) \qquad h_1(z) = s_1(z), \quad h_2(z) = s_2(z) \quad \text{for} \quad z \in (0, e^{-e}).$$

The properties (i)–(vi) of Lemma 3.2(b) now partially establish the result (c) of the theorem, the remaining step being to extend the continuity $h_1(z)$ and $h_2(z)$ from $(0, e^{-e}]$ to the closed interval $[0, e^{-e}]$. But this follows immediately by noting that

$$h_1(0) = 0 = \lim_{z \to 0} s_1(z) = \lim_{z \to 0} h_1(z),$$

and

$$h_2(0) = 1 = \lim_{z \to 0} s_2(z) = \lim_{z \to 0} h_2(z),$$

where we have used the convention (2.1), together with (3.2) and (4.7). Q.E.D.

References

1. Garrett Birkhoff and Gian-Carlo Rota, Ordinary Differential Equations, 3rd ed., Wiley, NY, 1978.
2. R. Arthur Knoebel, Exponentials reiterated, this MONTHLY, 88 (1981) 235–252.
3. David V. Widder, Advanced Calculus, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 1961.

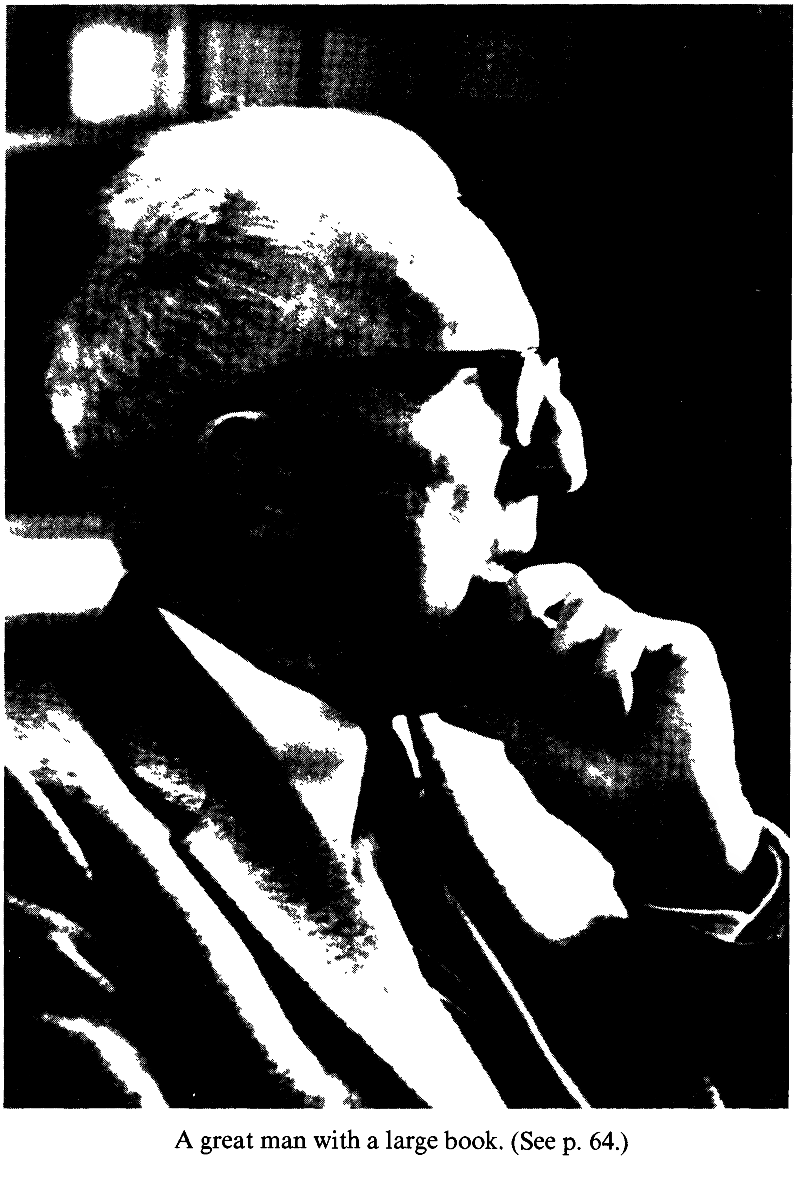
---

159. **MISCELLANEA**

Because other planets might be physically very different from ours, scientists there might use mathematics... very unlike ours... their 'geometry' could be something rather strange, largely topological, say, and geared to flexible structures rather than fixed sizes or shapes.

Just like some of our mathematics by the sound of it.

—Review of *The Limits of Science* by Nicholas Rescher, University of California Press (1985), by David Miller in *Nature* 314 (1985) 684.



A great man with a large book. (See p. 64.)

Lattice Points and the Taxicab Metric

E 2989 [1983, 212]. *Proposed by M. Goldberg, University of Waterloo, and S. C. Locke, Florida Atlantic University.*

Let $A = (a, a')$ and $B = (b, b')$ be lattice points (points in the plane with integral coordinates), and let $d(A, B)$ denote $|a - b| + |a' - b'|$. Let S be the set of points at d-distance at most k from the origin. Calculate

$$f(k) = \sum_{A, B \in S} d(A, B).$$

Solution by William A. Newcomb, Lawrence Livermore National Laboratory.

$$f(k) = \frac{2}{15}k(k+1)(2k+1)(7k^2+7k+6),$$

if the sum is understood to be over all *unordered* pairs of points A, B in S.

Proof. Horizontal and vertical segments contribute equally to the sum. Therefore,

$$f(k) = \sum_{-k \leqslant a < b \leqslant k} 2N_a N_b (b - a),$$

where N_x, for any integer x in the specified range, denotes the number of points in S with the horizontal coordinate x, namely, $2k + 1 - 2|x|$. We break up the sum as follows into three partial sums:

$$\sum_{0<a<b} + \sum_{a<b<0} = 2\sum_{0<a<b} 2N_a N_b (b-a),$$

$$\sum_{a=0<b} + \sum_{a<b=0} = 2\sum_{b>0} 2bN_0 N_b,$$

$$\sum_{a<0<b} 2N_a N_b (b-a) = \sum_{0<a,b} 2N_a N_b (b+a)$$

$$= 2\sum_{0<a<b} 2N_a N_b (b+a) + \sum_{b>0} 4bN_b^2.$$

The grand total is

$$\sum_{1}^{k} 4bN_b(N_0 + 2N_1 + \cdots + 2N_{b-1} + N_b),$$

wherein $N_b = 2k + 1 - 2b$, $N_0 + 2N_1 + \cdots = 2b(2k + 1 - b)$. The result, then, is

$$f(k) = 8(2k+1)^2 S_2 - 24(2k+1)S_3 + 16S_4,$$

where $S_n = 1^n + 2^n + \cdots + k^n$. Finally, by using known formulas for the power sums, one reduces this to the polynomial given above.

Also solved by J. Dou (Spain), M. Golomb, D. Hamlin, H. Honkasalo (Finland), J. B. M. Melissen (The Netherlands), W. Mixon, R. B. Nelsen, T. T. Nguyen, G. Sylvester, University of Arizona Problem Solving Group, and the proposers.

ANSWER TO PHOTO ON PAGE 24

Marston Morse, famous for, among other things, his book *The Calculus of Variations in the Large*.

A SIMPLE DERIVATION OF STIRLING'S ASYMPTOTIC SERIES

VICTOR NAMIAS
Department of Chemistry and Physics, Purdue University Calumet, Hammond, IN 46323

1. Introduction. Stirling's formula which approximates the factorial function plays a crucial role in probability theory and in statistical physics. In most physical applications, the simpler form

$$n! \simeq n^n e^{-n} \tag{1}$$

is quite sufficient when n is large.

From a mathematical point of view, the more accurate expression

$$n! = n^n e^{-n}\sqrt{2\pi n}\,\exp\left(\frac{1}{12n} - \frac{1}{360n^3} + \frac{1}{1260n^5} + \cdots\right) \tag{2}$$

credited to Stirling, but apparently found by de Moivre, does spectacularly better. Stirling's formula and its different versions have a fascinating history. In a recent issue of the American Journal of Physics **[4]**, Ian Tweddle of the University of Stirling, Stirling, Scotland, U.K., gives an interesting brief review of the historical facts surrounding the discovery of Stirling's formula.

As is well known, expansion (2) is an example of a diverging asymptotic series whose performance deteriorates as the number of terms is increased beyond a certain value. The derivation of formula (1) is quite simple and usually based on approximating $\ln n! = \sum_{k=1}^{n} \ln k$ by an integral.

The proof of expansion (2) is somewhat more elaborate and can be found in more advanced texts in mathematical analysis. Most derivations are directly based on the Euler-Maclaurin formula, whose demonstration, although not difficult, still lies beyond the freshman college level. In H.& B. Jeffreys' Mathematical Physics **[2]**, for example, the demonstration takes up four large pages of text. Attempts have been made to find a simpler derivation of Stirling's series. Recently, N. D. Mermin **[3]** presented a proof which started off simply and clearly but very quickly got into muddier waters. Not only does it become very laborious to calculate the higher power terms of the series, but there seems to be no systematic way of obtaining them. In contrast, the demonstration we propose is based on the well-known Legendre duplication formula for the factorial function and only calls for a relatively elementary level of mathematical sophistication. In fact, when the calculations in Section 2 are carried out term by term, i.e., by writing out the first few terms in the expansion of some elementary functions and identifying the coefficients of the series instead of using the general summation formalism, the results can be obtained very simply and with a minimum of intellectual effort. When the method is applied in its full generality and the results are compared with those of the standard procedure involving Bernoulli numbers, a recurrence relation for the Bernoulli numbers is obtained. The method can also be applied to the triplication formula or, in general, to the Gauss's multiplication formula, and a whole family of different recurrence relations for the Bernoulli numbers can be derived.

2. Derivation based on the duplication formula. Using the basic property of the factorial function: $n! = n(n-1)!$ or, alternatively, of the gamma function: $\Gamma(n+1) = n\Gamma(n)$, we can write

$$n\Gamma(n) = n(n-1)(n-2)(n-3)\cdots 1$$

Victor Namias received his M.S. in electrical and nuclear engineering from the Ecole Polytechnique of the Université Libre de Bruxelles, Belgium. He began employment as a nuclear engineer and, simultaneously, he pursued advanced studies in the field of Physics at the Carnegie Institute of Technology (presently, Carnegie-Mellon University), where he earned an M.S. and a Ph.D. in Physics. Although inclined by profession to the applied side of mathematics, he is also quite sensitive to the intrinsic beauty of pure mathematics, and he also has a passionate interest in music and musicology.

and

$$\Gamma\left(n+\frac{1}{2}\right)=\left(n-\frac{1}{2}\right)\left(n-\frac{3}{2}\right)\left(n-\frac{5}{2}\right)\cdots\left[n-\frac{(2n-1)}{2}\right]\Gamma\left[n-\frac{(2n-1)}{2}\right].$$

Multiplying these two expressions by one another yields

$$\begin{aligned}(3)\qquad n\Gamma(n)\Gamma\left(n+\frac{1}{2}\right)&=n\left(\frac{2n-1}{2}\right)(n-1)\left(\frac{2n-3}{2}\right)(n-2)\\&\quad\times\cdots\left[n-\frac{(2n-1)}{2}\right]\Gamma\left[n-\frac{(2n-1)}{2}\right]\\&=\frac{2n}{2}\left(\frac{2n-1}{2}\right)\left(\frac{2n-2}{2}\right)\left(\frac{2n-3}{2}\right)\left(\frac{2n-4}{2}\right)\\&\quad\times\cdots\left[\frac{2n-(2n-1)}{2}\right]\Gamma\left[n-\frac{(2n-1)}{2}\right]\\&=\frac{2n\Gamma(2n)}{2^{2n}}\Gamma\left(\frac{1}{2}\right).\end{aligned}$$

Here, $\Gamma(1/2)$ can be considered as a constant to be determined. Anyone familiar with the generalized definition of the gamma function through Euler's integral of course knows that $\Gamma(1/2)=\sqrt{\pi}$. Equation (3) can be written

$$(4)\qquad \Gamma(2n)=\frac{1}{\sqrt{\pi}}\Gamma(n)\Gamma\left(n+\frac{1}{2}\right)2^{2n-1},$$

and is the well-known Legendre duplication formula.

We now seek to improve on the crude Stirling formula (1) by writing

$$(5)\qquad \Gamma(n)=n^{n-1}e^{-n}G(n),$$

where $G(n)$ is a function to be determined. Inserting this expression into the duplication formula (4), we find at once

$$\frac{G(2n)}{G(n)G\left(n-\frac{1}{2}\right)}=\frac{\sqrt{e}}{\sqrt{\pi}}\left(1-\frac{1}{2n}\right)^{n}\frac{1}{\sqrt{n-\frac{1}{2}}}=\sqrt{e}\left(1-\frac{1}{2n}\right)^{n}\frac{1}{\sqrt{2\pi}}\frac{\sqrt{2n}}{\sqrt{n}\sqrt{n-\frac{1}{2}}}.$$

This form suggests that further simplification can be achieved by letting

$$(6)\qquad G(n)=F(n)\sqrt{2\pi n}.$$

We find, in terms of F,

$$(7)\qquad \frac{F(2n)}{F(n)F\left(n-\frac{1}{2}\right)}=\sqrt{e}\left(1-\frac{1}{2n}\right)^{n}=\exp\left[\frac{1}{2}+n\ln\left(1-\frac{1}{2n}\right)\right].$$

From the expansion

$$\ln(1-x)=-\sum_{s=1}^{\infty}\frac{x^s}{s},$$

the right-hand side of equation (7) becomes

$$\exp\left[\frac{1}{2}-\sum_{s=1}^{\infty}\frac{1}{2^s n^{s-1}s}\right].$$

Letting $k=s-1$, we obtain

$$\frac{F(2n)}{F(n)F\left(n-\frac{1}{2}\right)} = \exp\left[-\sum_{k=1}^{\infty}\frac{1}{2^{k+1}n^k(k+1)}\right]. \tag{8}$$

Equation (8) further suggests that $F(n)$ might also be written as an exponential

$$F(n) = \exp\left(\sum_{k=1}^{\infty}\frac{a_k}{n^k}\right), \tag{9}$$

where the a_k are a set of constant coefficients to be determined.

Stirling's formula (5) then becomes the series

$$\Gamma(n) = n^{n-1}e^{-n}\sqrt{2\pi n}\exp\left(\sum_{k=1}^{\infty}\frac{a_k}{n^k}\right). \tag{10}$$

Inserting equation (9) into the left-hand side of (8) yields

$$\exp\left[\sum_{k=1}^{\infty}\frac{a_k}{n^k}\left(\frac{1}{2^k}-1\right)-\sum_{k=1}^{\infty}\frac{a_k}{\left(1-\frac{1}{2n}\right)^k n^k}\right]. \tag{11}$$

We now use the expansion

$$\frac{1}{1-x} = \sum_{s=0}^{\infty}x^s,$$

valid for $x < 1$, and by successive differentiations, we obtain

$$\frac{1}{(1-x)^m} = \sum_{s=0}^{\infty}\frac{s!x^{s-m+1}}{(s-m+1)!(m-1)!}. \tag{12}$$

Letting $x = 1/2n$, and inserting into expression (11), we find

$$\exp\left[\sum_{k=1}^{\infty}\left(\frac{1}{2^k}-1\right)\frac{a_k}{n^k}-\sum_{m=1}^{\infty}\sum_{s=0}^{\infty}\frac{s!a_m}{2^{s-m+1}n^{s+1}(s-m+1)!(m-1)!}\right]. \tag{13}$$

We now let $k = s+1$ in the double summation and obtain

$$\exp\left[\sum_{k=1}^{\infty}\left(\frac{1}{2^k}-1\right)\frac{a_k}{n^k}-\sum_{m=1}^{\infty}\sum_{k=1}^{\infty}\frac{(k-1)!a_m}{2^{k-m}n^k(k-m)!(m-1)!}\right]. \tag{14}$$

We note that the m summation need not go to infinity because for $m > k$ the denominator in the double summation becomes infinite and makes no contribution to the sum. Thus, expression (14) can also be written

$$\exp\left\{\sum_{k=1}^{\infty}\left[\left(\frac{1}{2^k}-1\right)a_k-(k-1)!\sum_{m=1}^{k}\frac{a_m}{2^{k-m}(k-m)!(m-1)!}\right]\frac{1}{n^k}\right\}. \tag{15}$$

We now identify expressions (8) and (15) by equating the coefficients of $1/n^k$ in both expansions. The result is

$$\frac{1}{2(k+1)} = (2^k-1)a_k+(k-1)!\sum_{m=1}^{k}\frac{a_m}{2^{-m}(k-m)!(m-1)!}. \tag{16}$$

If we exclude the a_k term in the summation and solve equation (16) for a_k, we obtain a recurrence relation for these coefficients:

$$a_k = \frac{1}{1 - 2^{k+1}}\left[-\frac{1}{2(1+k)} + (k-1)!\sum_{m=1}^{k-1}\frac{2^m a_m}{(k-m)!(m-1)!}\right]. \tag{17}$$

From equation (16) or (17), we find successively:

$$a_1 = 1/12; \quad a_2 = 0; \quad a_3 = -1/360; \quad a_4 = 0; \quad a_5 = 1/1260; \text{ etc.},$$

so that Stirling's asymptotic series can be calculated easily and in a systematic way.

3. Recurrence Relations for the Bernoulli numbers. In the standard derivation of Stirling's asymptotic expansion [**1**], [**2**], the coefficients are found to be expressible in terms of the Bernoulli numbers B_m. We have

$$\Gamma(n) = n^{n-1}e^{-n}\sqrt{2\pi n}\exp\left(\sum_{m=2}^{\infty}\frac{B_m}{m(m-1)n^{m-1}}\right). \tag{18}$$

Comparing equation (10) with (18) reveals that the a_k coefficients and the Bernoulli numbers are connected by the simple relation

$$a_k = \frac{B_{k+1}}{k(k+1)}. \tag{19}$$

Relation (17) then yields a corresponding recurrence relation for the Bernoulli numbers. We find at once

$$B_k = \frac{1}{2(2^k - 1)}\left[(k-1) - k!\sum_{m=2}^{k-1}\frac{2^m B_m}{(k-m)!m!}\right]. \tag{20}$$

The above relation can also be written

$$B_k = \frac{-1}{2(2^k-1)}\sum_{m=0}^{k-1}\frac{k!2^m B_m}{(k-m)!m!} = \frac{1}{2(1-2^k)}\sum_{m=0}^{k-1}2^m C_k^m B_m, \tag{21}$$

and is valid for $k \geqslant 1$, with B_0 taken equal to 1. It is another addition to the arsenal of recurrence relations for the Bernoulli numbers, the most familiar of which is

$$\sum_{m=0}^{k-1} C_k^m B_m = 0 \quad \text{when} \quad k > 1, \quad \text{and} \quad \sum_{m=0}^{k-1} C_k^m B_m = 1 \quad \text{when} \quad k = 1.$$

The same procedure can also be applied to the more general Gauss's multiplication formula. For example, starting with the triplication formula

$$\Gamma(3n) = \frac{1}{2\pi}3^{3n-1/2}\Gamma(n)\Gamma\left(n+\frac{1}{3}\right)\Gamma\left(n+\frac{2}{3}\right), \tag{22}$$

and letting

$$\Gamma(n) = n^{n-1}e^{-n}\sqrt{2\pi n}\,F(n), \tag{23}$$

we find the following functional equation:

$$\frac{F(3n)}{F(n)F\left(n-\frac{1}{3}\right)F\left(n-\frac{2}{3}\right)} = e\left(1-\frac{1}{3n}\right)^{n+1/6}\left(1-\frac{2}{3n}\right)^{n-1/6} \tag{24}$$

Proceeding through the same steps as before, we obtain a new recurrence relation for the a_k coefficients:

$$a_k = \frac{1}{1-3^{k+1}}\left[\frac{(1-3k)2^k - (1+3k)}{6k(k+1)} + (k-1)!\sum_{m=1}^{k-1}\frac{(2^{k-m}+1)3^m a_m}{(m-1)!(k-m)!}\right]. \tag{25}$$

When the a_k coefficients are calculated through equation (25), they are found to be exactly the same as those obtained through equation (17) derived from the duplication formula. Thus, the exact same Stirling series is obtained. Once again, therefore, these coefficients are connected to the Bernoulli numbers and equation (25) translates into another recurrence relation for the Bernoulli numbers. Using equation (19), we find

$$B_k = \frac{1}{3(1-3^k)}\left[\left(2-\frac{3k}{2}\right)2^{k-1} - \left(\frac{3k}{2}-1\right) + k!\sum_{m=2}^{k-1}\frac{3^m(1+2^{k-m})B_m}{m!(k-m)!}\right]. \tag{26}$$

This result can be written in the more compact form

$$B_k = \frac{1}{3(1-3^k)}\sum_{m=0}^{k-1} 3^m(1+2^{k-m})C_k^m B_m. \tag{27}$$

Apparently, an infinite number of such recurrence relations can be generated starting from Gauss's multiplication formula. It is conjectured that in all cases the same Stirling series is obtained even though the recurrence relation for the coefficients of the series is different. In consequence, when the connection is made with the Bernoulli numbers, a whole family of recurrence relations such as (21) and (27) can be obtained.

References

1. M. Abramowitz and I. A. Stegun, Handbook of Mathematical Functions, National Bureau of Standards, Appl. Math. Series 55, 1965.
2. Sir H. Jeffreys and B. Swirles (Lady Jeffreys), Methods of Mathematical Physics, Cambridge University Press, 1962.
3. N. D. Mermin, Stirling's formula!, Amer. J. Phys., 52 (1984) 362.
4. I. Tweddle, Approximating $n!$. Historical origins and error analysis, Amer. J. Phys., 52 (1984) 487.

21 − 6 = 15: A CONNECTION BETWEEN TWO DISTINGUISHED GEOMETRIES

ALBRECHT BEUTELSPACHER
Fachbereich Mathematik der Universität, Saarstr. 21, D-6500 Mainz, Federal Republic of Germany

Two of the most remarkable geometric individuals are the projective plane **P** of order 4 and the 3-dimensional projective space Π of order 2. In this paper we shall study the mutual relations between these gems: Π will be obtained as **P** minus a hyperoval.

First, we shall prove the uniqueness and existence of the projective plane of order 4 in an extremely easy combinatorial manner. We shall also determine the order of the full collineation group of **P**. After having constructed the projective space Π out of **P**, we will be able to see some remarkable objects of Π (such as ovoids, reguli and spreads) in **P**. As a link between our geometries **P** and Π we shall use the complete graph K_6 on six vertices. It is surprising that we can describe all points, lines and planes of **P** and Π by means of a structure as simple as K_6.

Albrecht Beutelspacher: I was educated at the Universities of Tübingen and Mainz, F.R.G., where I received my Ph.D. under the supervision of Judita Cofman in 1976 and took my "Habilitation" in 1980. Since then I have taught at the University of Mainz and have visited several Universities in Europe and Canada. A two-volume book *Einführung in die endliche Geometrie* appeared in 1982/83; presently I am writing a book on recreational mathematics.

My mathematical interests include combinatorics, finite geometry and coding theory. Apart from my family my main nonmathematical interests are books and music; I play the organ regularly.

We do not claim that there are any new facts in this paper—the interested reader should consult in particular [1], [5], [9]. In [7, 14.3] some results of our Sections 1 and 2 can be found. Also the papers [2], [3], [4], [12], [13] deal with questions very close to those discussed here.

1. Quadrangles, subplanes and hyperovals in a projective plane of order 4. A *projective plane* is an incidence structure $\mathbf{P} = (\mathscr{P}, \mathscr{L}, I)$. The elements of $\mathscr{P}$ are called *points*, the elements of $\mathscr{L}$ lines, and I is an *incidence relation* for which the following axioms hold:

(1) Any two distinct points P and Q of $\mathbf{P}$ are incident with exactly one line; this uniquely determined line is denoted by PQ.

(2) Any two distinct lines l and m of $\mathbf{P}$ have a (unique) point of intersection; we call this point $l \cap m$.

(3) There exist four points, no three of which are collinear (i.e., no three of which are incident with a common line of $\mathbf{P}$).

In this paper we shall consider only *finite* projective planes, i.e., projective planes with a finite number of points. It is easy to show (see for example [8]) that for any finite projective plane $\mathbf{P}$ there exists a positive integer n (called the *order* of $\mathbf{P}$) such that the following holds:

(a) Any line of $\mathbf{P}$ has exactly $n + 1$ points;
(b) any point is incident with exactly $n + 1$ lines;
(c) the number of points and the number of lines of $\mathbf{P}$ are both equal to $n^2 + n + 1$;
(d) $n \geqslant 2$.

Fig. 1 is the famous visualization of the projective plane of order 2. Two remarks are in order. (1) In this picture, we have a "line" which is not straight. This happens frequently in finite geometry; in fact, there is no finite projective plane which can be drawn in the euclidean plane with straight lines only. (2) We speak of *the* projective plane of order 2. In this paper we shall show that there is a unique projective plane of order 4. After having read Section 2, it should be an easy exercise for the reader to prove the uniqueness of a projective plane of order 2.

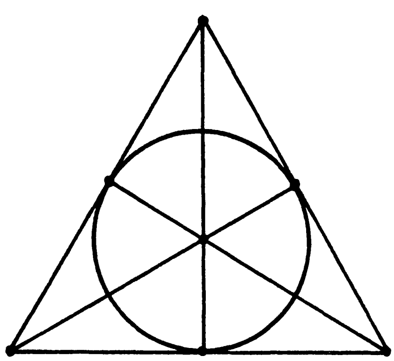


FIG. 1

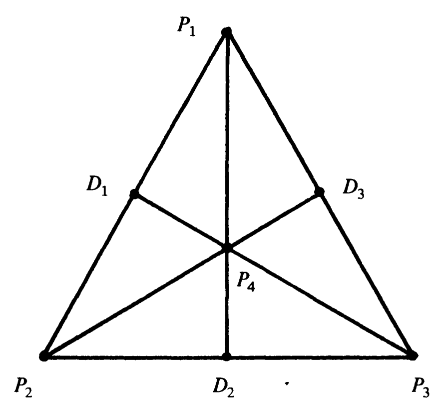


FIG. 2

In the remainder of this section we shall deal with an arbitrary projective plane $\mathbf{P}$ of order 4. Thus $\mathbf{P}$ has 21 points, 21 lines, 5 points on any line and 5 lines through any point.

A set of four points of $\mathbf{P}$, no three of which are collinear, is called a *quadrangle* of $\mathbf{P}$.

Denote by $\mathscr{Q} = \{P_1, P_2, P_3, P_4\}$ a quadrangle of $\mathbf{P}$. The lines P_iP_j $(1 \leqslant i, j \leqslant 4, i \neq j)$ are said to be the *diagonals* of $\mathscr{Q}$, and the points

$$D_1 = P_1P_2 \cap P_3P_4, \quad D_2 = P_1P_3 \cap P_2P_4 \quad \text{and} \quad D_3 = P_1P_4 \cap P_2P_3$$

are called the *diagonal points* of $\mathscr{Q}$. (Cf. Fig. 2.) The assertion of the following fundamental

Lemma may surprise the reader, but he should look back to Fig. 1.

1.1 LEMMA. *Any quadrangle of* **P** *has collinear diagonal points.*

Proof. Assume to the contrary that D_1, D_2, D_3 are not on a common line. Then there are precisely 9 lines which intersect $\bar{\mathcal{Q}} = \mathcal{Q} \cup \{D_1, D_2, D_3\}$ in at least two points (namely the 6 diagonals and the 3 lines incident with two diagonal points). Moreover, there exist exactly 11 lines having just one point in common with $\bar{\mathcal{Q}}$. (Through P_i there are two such lines and through D_j only one.)

Since $9 + 11 < 21$, there is a line x of **P** having no point in common with $\bar{\mathcal{Q}}$. Therefore, x intersects the 6 diagonals of $\mathcal{Q}$ in 6 distinct points, a contradiction, since any line of **P** has only 5 points. □

An alternative formulation of Lemma 1.1 is *for any quadrangle* $\mathcal{Q}$ *of* **P**, $\bar{\mathcal{Q}}$ *is the point set of a subplane of order* 2 *of* **P**.

1.2 COROLLARY. *Denote by* $\mathcal{Q}$ *a quadrangle of* **P**. *Then there are exactly two points* P *and* Q *of* **P** *which are not incident with any diagonal of* $\mathcal{Q}$. *Moreover,* $\mathcal{Q}^* = \mathcal{Q} \cup \{P, Q\}$ *is a set of* 6 *points, no three of which are collinear.*

Proof. The 6 diagonals of $\mathcal{Q}$ cover $5 + 4 + 3 + 2 + 3 + 2 = 19$ points of **P**. The remaining 2 points are the points on the line d through D_1, D_2, D_3 different from D_1, D_2, D_3. □

In a projective plane of order n, a set of $n + 2$ points, no three of which are collinear, is said to be a *hyperoval*. Thus 1.2 says that any quadrangle $\mathcal{Q}$ of **P** is contained in a unique hyperoval $\mathcal{Q}^*$; the points of $\mathcal{Q}^* - \mathcal{Q}$ are the points outside $\bar{\mathcal{Q}}$ on the line d connecting the diagonal points of $\mathcal{Q}$.

We shall now consider a hyperoval $\mathcal{H}$ of **P**.

1.3 LEMMA. *Denote by* $\mathcal{H}$ *a hyperoval of* **P**. *Then any line of* **P** *intersects* $\mathcal{H}$ *in just* 2 *points or none.*

Proof. Consider a line l which has at least one point, say P, in common with $\mathcal{H}$. Since P is joined to all 5 points of $\mathcal{H} - \{P\}$ by a line and since P is incident with exactly 5 lines, any line through P has exactly 2 points in common with $\mathcal{H}$. This holds in particular for l. □

1.4 PROPOSITION. *Let* $\mathcal{H} = \{A, B, C, D, E, F\}$ *be a hyperoval of* **P**. *Then the lines* AB, CD, EF *pass through a common point of* **P**.

Proof. By 1.3, the line joining $AB \cap CD$ to E must contain F. □

We call the lines of **P** which intersect $\mathcal{H}$ in two points 3-*lines*; the other lines are said to be 5-*lines*.

1.5 PROPOSITION. (a) *Through any point of* **P** *outside* $\mathcal{H}$ *there are exactly two* 5-*lines.*
(b) *There exist exactly six* 5-*lines.*

Proof. (a) By 1.3, through any point outside $\mathcal{H}$ there are exactly three 3-lines.
(b) If b denotes the number of 5-lines, it follows by (a) that

$$b \cdot 5 = (21 - 6) \cdot 2. \square$$

One obtains the dual of a projective plane $\mathbf{P} = (\mathcal{P}, \mathcal{L}, I)$ by interchanging the rôles of points and lines. More precisely, the *dual* of **P**, denoted by $\mathbf{P}^d$, is the incidence structure $(\mathcal{L}, \mathcal{P}, I^d)$ for which

$$I^d = \{(l, P) \mid (P, l) \in I\}.$$

It is easy to verify that $\mathbf{P}^d$ is again a projective plane. Moreover, if **P** is finite and has order n, then also $\mathbf{P}^d$ has order n.

Now, 1.5 yields immediately

1.6 COROLLARY. *If $\mathscr{H}$ is a hyperoval of* $\mathbf{P}$, *then the set $\mathscr{k}$ of 5-lines is a hyperoval of* $\mathbf{P}^d$. □

A particularly interesting property of the projective plane of order 4 is that in $\mathbf{P}$ affine planes of order 3 are hidden (cf. [**11**]). An affine plane of order 3 is an incidence structure with 9 points and 12 lines such that any line has 3 points and any two distinct points are incident with a unique common line (see Fig. 3). Can we already "see" an affine plane of order 3 in $\mathbf{P}$? For that purpose, we consider a hyperoval $\mathscr{H}=\{A,B,C,D,E,F\}$ of $\mathbf{P}$. The three lines AB, AC, BC intersect within $\mathscr{H}$, so they cover disjointly a set $\mathscr{N}$ of 9 points outside $\mathscr{H}$.

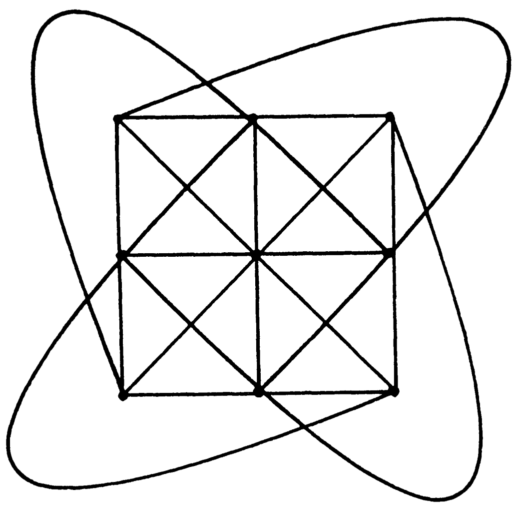


FIG. 3

The same is true for the lines DE, DF and EF. Moreover, since any of the first three lines intersects any of the latter lines, the lines DE, DF, EF cover also $\mathscr{N}$.

So we have already six lines, and the other six lines will be our 5-lines. Since any 5-line meets any 3-line in a point outside $\mathscr{H}$, any 5-line has exactly three points in common with $\mathscr{N}$.

It follows that $\mathscr{N}$ together with these 12 lines is an affine plane of order 3.

REMARK. Some authors call—with good reason—this affine plane in this situation a "hermitian curve" or a "unital of order 2" (see, e.g., [**8**], p. 246).

Another interesting property of $\mathbf{P}$ is that the theorems of Desargues and Pappus are valid. We begin with Desargues's theorem.

Two triangles A_1, A_2, A_3 and B_1, B_2, B_3 of a projective plane are said to be *perspective from a point*, if the lines A_iB_i $(1 \leqslant i \leqslant 3)$ pass through a common point $Z \neq A_i, B_i$. Dually, these triangles are called *perspective from a line*, if the points

$$P_{12} = A_1A_2 \cap B_1B_2, \quad P_{13} = A_1A_3 \cap B_1B_3 \quad \text{and} \quad P_{23} = A_2A_3 \cap B_2B_3$$

are on a common line. A projective plane is called *desarguesian* if any two triangles which are perspective from a point are perspective from a line (see Fig. 4).

1.7 THEOREM. *A projective plane* $\mathbf{P}$ *of order* 4 *is desarguesian.*

Proof. We handle only the interesting case in which the union of the two triangles A_1, A_2, A_3 and B_1, B_2, B_3 is a hyperoval of $\mathbf{P}$. Assume that these triangles are perspective from a point. Then the lines A_1A_2, A_1A_3, A_2A_3 and B_1B_2, B_1B_3, B_2B_3 form a parallel class in the affine plane $\mathbf{A}$ of order 3 constructed above, respectively. Looking back to Fig. 3 we see that the points P_{12}, P_{13}, P_{23} are the points of a line of $\mathbf{A}$. □

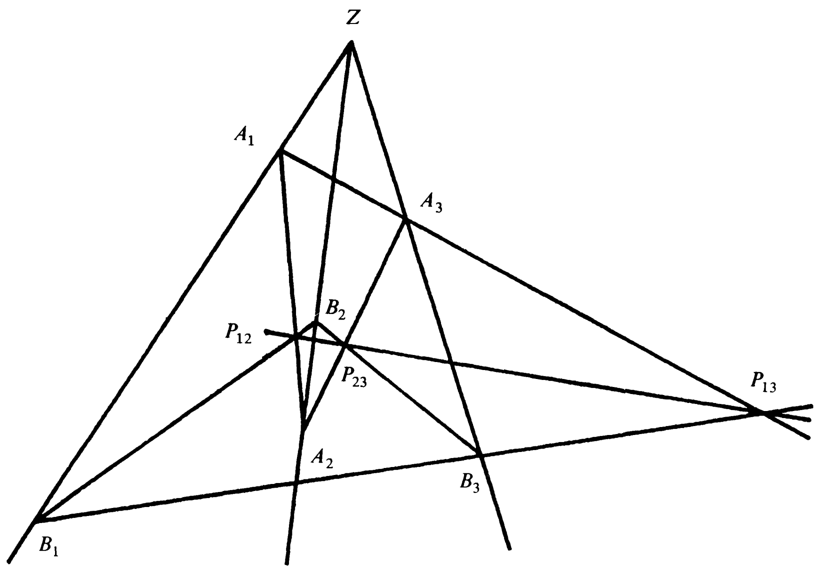


FIG. 4

We shall show that **P** is also pappian. We say that in a projective plane the *theorem of Pappus* is valid if the following holds:

For any two distinct lines l and m and any two triples (A_1, A_2, A_3), (B_1, B_2, B_3) of points with

$$A_i I l,\ B_i I m,\ A_i \neq l \cap m \neq B_i \quad (1 \leqslant i \leqslant 3),$$

the points

$$Q_{12} = A_1B_2 \cap B_1A_2, \quad Q_{13} = A_1B_3 \cap B_1A_3 \quad \text{and} \quad Q_{23} = A_2B_3 \cap B_2A_3$$

are collinear (see Fig. 5).

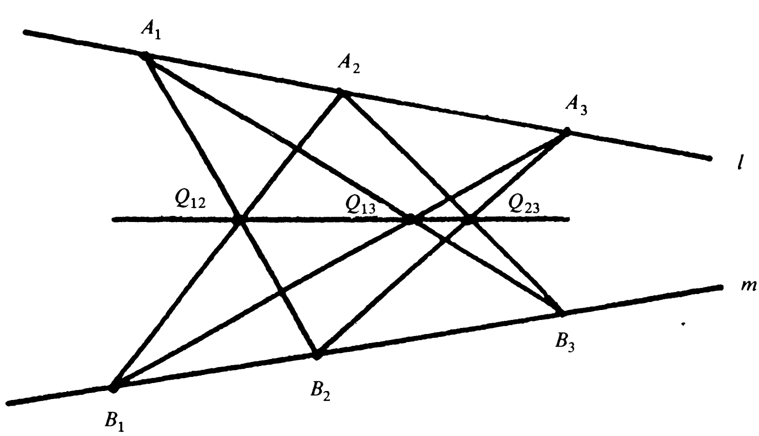


FIG. 5

1.8 THEOREM. *A projective plane* **P** *of order* 4 *is pappian, i.e., the theorem of Pappus holds in* **P**.

Proof. If $Q_{12} I A_3B_3$, $Q_{13} I A_2B_2$ and $Q_{23} I A_1B_1$, then it can be shown that the nine points $A_1, A_2, A_3, B_1, B_2, B_3, Q_{12}, Q_{13}, Q_{23}$ are the points of an affine subplane of order 3. (Cf. the proof of 1.1.) Hence Q_{12}, Q_{13}, Q_{23} are collinear.

If, for example, $Q_{12} \not{I} A_3 B_3$, then the lines $Q_{12}A_1$, $Q_{12}B_1$, $Q_{12}A_3$ and $Q_{12}B_3$ are four distinct lines through Q_{12}, none of which contains Q_{13} or Q_{23}. Hence these points must be incident with the fifth line through Q_{12}. □

2. Uniqueness and existence of a projective plane of order 4. In this section we shall prove that there exists a unique projective plane of order 4. In order to do this, we shall use the hyperovals in such planes. First, we list some graph-theoretic notions, all of which can be found in **[6]**. If $\mathcal{S}$ is a set of n elements, we denote by $\binom{\mathcal{S}}{2}$ the $\binom{n}{2}$ unordered pairs of $\mathcal{S}$. Then the incidence structure $K_n = (\mathcal{S}, \binom{\mathcal{S}}{2}, \in)$ is called the *complete graph* on *n vertices*; the elements of $\binom{\mathcal{S}}{2}$ are said to be the *edges* of K_n. A *factor* of K_n is a set $\mathcal{F}$ of edges such that any vertex is contained in exactly one edge of $\mathcal{F}$. Obviously, K_n has a factor if and only if n is an even number.

2.1 LEMMA. (a) *Any edge of K_6 is contained in precisely three factors.*
(b) *The graph K_6 has exactly* 15 *factors.*

Proof. (a)

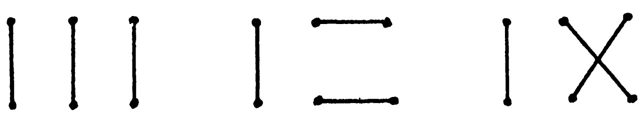


FIG. 6

(b) Denote by Φ the number of factors of K_6. Since K_6 has exactly $\binom{6}{2}$ edges, by (a) we have

$$\binom{6}{2} \cdot 3 = \phi \cdot 3. \square$$

A set $\mathscr{f}$ of factors of K_n is said to be a *factorization* of K_n if any edge is contained in exactly one factor of $\mathscr{f}$.

2.2 PROPOSITION. (a) *Any factor of K_6 is contained in precisely* 2 *factorizations*;
(b) *K_6 has exactly* 6 *factorizations*;
(c) *any two distinct factorizations of K_6 have at most one common factor.*

Proof. (a)

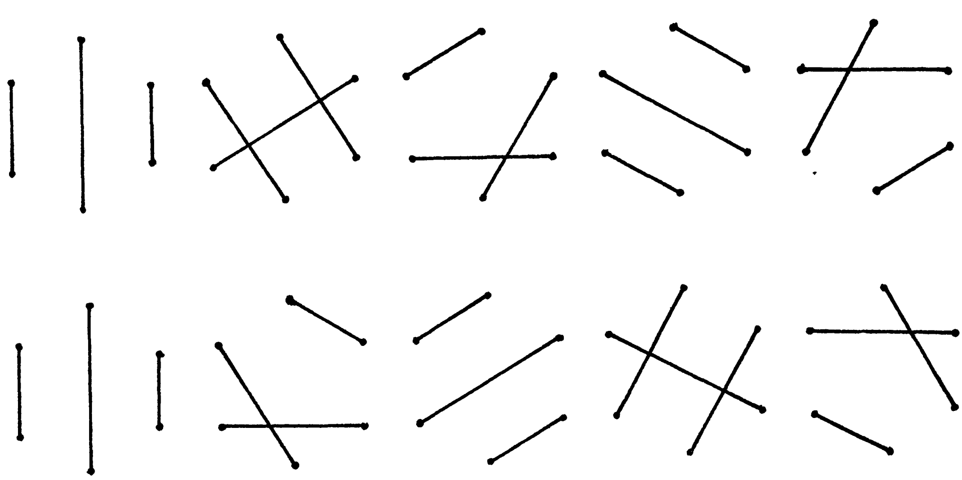


FIG. 7

(b) Denote by ϕ the number of factorizations of K_6. Since any factorization has exactly 5 factors, it follows in view of 2.1 and (a) that

$$\phi \cdot 5 = 15 \cdot 2.$$

(c) Follows by the proof of (a). □

At this stage, maybe the reader would like to ask the author: What on earth has this to do with the projective plane of order 4? Let's hope that the following Lemma will answer this question.

2.3 LEMMA. *Denote by* $\mathbf{P}$ *a projective plane of order* 4, *and let* $\mathscr{H}$ *be a hyperoval of* $\mathbf{P}$. *Denote by* K_6 *the complete graph whose vertices are the points of* $\mathscr{H}$, *and whose edges are (essentially) the* 3-*lines.*

(a) *If* P *is a point of* $\mathbf{P}$ *outside* $\mathscr{H}$, *then the* 3-*lines through* P *induce a factor* $\mathscr{F}(P)$ *of* K_6.

(b) *If* $\mathscr{F}$ *is a factor of* K_6, *then there is a point* P *with* $\mathscr{F} = \mathscr{F}(P)$.

(c) *If* l *is a* 5-*line, then* $\not{f}(l) = \{\mathscr{F}(P) | PIl\}$ *is a factorization of* K_6.

(d) *For any factorization* $\not{f}$ *of* K_6, *there is a* 5-*line* l *with* $\not{f} = \not{f}(l)$.

Proof. (a) follows from 1.3.

Since $21 - 6 = 15$, (b) follows from 2.1. (Another proof of this fact was given in 1.4)

(c) Any 3-line intersects l in a unique point.

(d) There are six 5-lines and six factorizations of K_6. □

2.4 THEOREM. *Denote by* $\mathbf{P}$ *and* $\mathbf{P}'$ *two projective planes of order* 4, *and let* $\mathscr{H}$ *and* $\mathscr{H}'$ *be hyperovals of* $\mathbf{P}$ *and* $\mathbf{P}'$, *respectively. Then any bijective map* φ *from* $\mathscr{H}$ *onto* $\mathscr{H}'$ *induces a unique isomorphism from* $\mathbf{P}$ *onto* $\mathbf{P}'$.

Proof. Clearly, φ maps factors onto factors. Therefore, in view of 2.3(b), φ induces in a unique way a bijection from the points of $\mathbf{P}$ onto the points of $\mathbf{P}'$. We have to show that this bijection (also called φ) is an isomorphism. Since φ is a bijection, it maps 3-lines onto 3-lines. Since φ maps factorizations onto factorizations, it maps 5-lines onto 5-lines. Finally, since φ preserves incidence, it is an isomorphism. □

As a corollary we have

2.5 THEOREM. (a) *Up to isomorphism, there is at most one projective plane of order* 4.

(b) *The automorphism group of a projective plane of order* 4 *has a subgroup isomorphic to* $\mathscr{S}_6$, *the symmetric group on* 6 *symbols.*

Proof. Since any projective plane of order 4 has a hyperoval (see 1.2), (a) follows from 2.4.

(b) Let $\mathbf{P}$ be a projective plane of order 4, and denote by $\mathscr{H}$ a hyperoval of $\mathbf{P}$. By 2.4, any permutation of $\mathscr{H}$ is induced by an automorphism of $\mathbf{P}$. □

Now we have to prove the existence of a projective plane of order 4.

2.6 THEOREM. *There exists a projective plane of order* 4.

Proof. Denote by $K_6 = (\mathscr{S}, \binom{\mathscr{S}}{2}, \in)$ the complete graph on 6 vertices. We define the incidence structure $\mathbf{P} = (\mathscr{P}, \mathscr{L}, I)$ as follows:

$$\mathscr{P} = \mathscr{S} \cup \{\mathscr{F} | \mathscr{F} \text{ is a factor of } K_6\},$$

$$\mathscr{L} = \binom{\mathscr{S}}{2} \cup \{\not{f} | \not{f} \text{ is a factorization of } K_6\}.$$

The incidence relation I will be defined in the following way:

$$VIe \Leftrightarrow V \in e \quad \left(\text{for } V \in \mathscr{S}, e \in \binom{\mathscr{S}}{2}\right),$$

$$V I f \qquad \text{(for } V \in \mathcal{S},\ f \text{ a factorization)},$$
$$\mathcal{F} I e \Leftrightarrow e \in \mathcal{F} \quad \left(\text{for } e \in \binom{\mathcal{S}}{2},\ \mathcal{F} \text{ a factor}\right),$$
$$\mathcal{F} I f \Leftrightarrow \mathcal{F} \in f \quad (\text{for a factor } \mathcal{F} \textit{ and a factorization } f).$$

We claim that **P** is a projective plane of order 4. Since **P** has 21 (= 6 + 15) points, we have only to show that **P** is a projective plane.

Step 1. Any two distinct points of **P** are incident with exactly one line of **P**. The only possibility that does not follow immediately from the definition is when the two points under consideration are two distinct factors $\mathcal{F}$ and $\mathcal{F}'$. By 2.1(a), $\mathcal{F}$ and $\mathcal{F}'$ have at most one edge in common. If $\mathcal{F}$ and $\mathcal{F}'$ intersect in an edge e, then e is the unique line of **P** incident with $\mathcal{F}$ and $\mathcal{F}'$. Suppose now $\mathcal{F} \cap \mathcal{F}' = \varnothing$. By 2.1, there are exactly 8 $(= 15 - 1 - 3(3 - 1))$ factors disjoint from $\mathcal{F}$. So, by 2.2(a) and (c), any factor disjoint from $\mathcal{F}$ is contained in exactly one of the two factorizations through $\mathcal{F}$. This holds in particular for $\mathcal{F}'$.

Step 2. Any two distinct lines of **P** are incident with precisely one common point.

First, two distinct edges either intersect in a vertex or are contained in a (unique) factor. Next, if e is an edge and f a factorization, then, by definition, there is a unique factor $\mathcal{F}$ with $e \in \mathcal{F} \in f$. Finally, consider two distinct factorizations f and f'. Any of the 5 factors of f lies in a unique factorization different from f. Since by 2.2(c) any two factorizations intersect in at most one factor, there are 5 factorizations which have exactly one factor in common with f. Thus, f and f' intersect in a unique common factor.

Of course, **P** has a quadrangle (for example, take four points of $\mathcal{S}$). Hence **P** is a projective plane. □

3. The collineation group of P. Since an automorphism of a projective plane preserves collinearity, geometers generally prefer to call it a *collineation*. In this section, we shall determine the order of the full collineation group of **P**, the projective plane of order 4, and show that there exists a subgroup that is sharply transitive on the ordered quadrangles of **P**.

3.1 LEMMA. (a) *In* **P** *there exist exactly* $21 \cdot 20 \cdot 16 \cdot 9$ *ordered quadrangles* $\mathcal{Q} = (A, B, C, D)$.
(b) *In* **P** *there exist exactly* $21 \cdot 20 \cdot 16 \cdot 9 \cdot 2 = 120960$ *ordered hyperovals* $\mathcal{H} = (A, B, C, D, E, F)$.

Proof. (a) For the first point A of $\mathcal{Q}$ we have 21 possibilities, for B we have 20. In order to choose the third point C, any point off the line AB is possible. Finally, the last point D can be chosen arbitrarily outside the 3 lines AB, AC and BC. Since these lines cover exactly $5 + 4 + 3 = 12$ points, the assertion follows.

(b) There are $21 \cdot 20 \cdot 16 \cdot 9$ ways to choose the ordered quadrangle (A, B, C, D); then, by 1.2, the points D and E are uniquely determined—up to order. There are 2 ways to order these remaining points. □

3.2 THEOREM. *The group* Γ *of all collineations of* **P** *has exactly* 120960 *elements. For any two ordered hyperovals* $(A_1, \ldots, A_6)$ *and* $(B_1, \ldots, B_6)$ *there is precisely one collineation which maps* A_i *onto* B_i $(1 \leqslant i \leqslant 6)$.

Proof. From the definition, any collineation of **P** maps an ordered hyperoval onto an ordered hyperoval. Hence the assertion follows by 2.4 and 3.1(b). □

3.3 LEMMA. *Denote by* $\mathcal{Q} = \{P_1, P_2, P_3, P_4\}$ *a quadrangle of* **P**, *and let* φ *be a collineation of* **P** *with* $\varphi(P_i) = P_i$ *for* $1 \leqslant i \leqslant 4$. *Then there are two possibilities: Either* (a) φ *is the identity, or* (b) *the fixed points of* φ *are exactly the* 7 *points of* $\bar{\mathcal{Q}}$, φ *is a product of* 7 *disjoint transpositions, and* (*consequently*) φ *is an odd permutation.*

Proof. Clearly, φ fixes the 6 diagonals and hence the diagonal points of $\mathcal{Q}$. So, any point of $\bar{\mathcal{Q}}$

is a fixed point of φ.

Suppose that φ fixes a point X outside $\bar{\mathscr{Q}}$. There exists a line l of $\bar{\mathscr{Q}}$ (i.e., a diagonal or the line through the diagonal points) through X. Since l is incident with at least 4 fixed points, all 5 points of l are fixed under φ. Now, let X' be an arbitrary point outside l and $\bar{\mathscr{Q}}$. Denote by l' the line of $\tilde{\mathscr{Q}}$ through X'. We have already observed that l' is fixed under φ. There exists a point P of $\bar{\mathscr{Q}}$ which is neither on l nor on l'. Then the line $l'' = PX'$ intersects l in a point $\mathscr{Q}$ ($\neq P$), which is a fixed point. Consequently, l'' is a fixed line of φ. Hence $X' = l' \cap l''$ is fixed by φ.

Thus we have proved that either φ is the identity, or no point outside $\bar{\varphi}$ is fixed by φ. Since any orbit of φ has length 1 or 2 and there are exactly 14 points outside $\bar{\mathscr{Q}}$, our assertion is proved. □

3.4 COROLLARY. *Denote by φ a collineation of* **P**. *If φ is an even permutation and fixes a quadrangle pointwise, then φ is the identity. There exists a collineation which is an odd permutation.*

Proof. If a nontrivial collineation φ fixes the points of a quadrangle, then, by 3.2, φ is the product of an odd number of transpositions, hence an odd permutation.

Let $\mathscr{H} = \{A, B, C, D, E, F\}$ be a hyperoval of **P** and define φ on $\mathscr{H}$ to be the transposition $(E\ F)$. Then φ fixes the ordered quadrangle (A, B, C, D) but is not the identity. □

3.5 THEOREM. *Denote by G the subgroup of Γ which consists of those collineations of* **P** *which are even permutations. Then G is sharply transitive on the ordered quadrangles of* **P**. *That is, for any two ordered quadrangles (P_1, P_2, P_3, P_4) and (Q_1, Q_2, Q_3, Q_4) of* **P**, *there is precisely one collineation φ of G with $\varphi(P_i) = Q_i$ $(1 \leqslant i \leqslant 4)$.*

Proof. By 3.4, the identity is the only element in G which fixes the points of an ordered quadrangle. In view of 3.4 and 3.2 it follows that

$$|G| = |\Gamma|/2 = 21 \cdot 20 \cdot 16 \cdot 9.$$

Since this is the number of ordered quadrangles (see 3.1), the assertion is proved. □

REMARK. The groups Γ and G are usually called PΓL(3, 4) and PGL(3, 4), respectively. With the theorem above it is now possible to prove the existence of the large Mathieu groups and the corresponding Steiner systems in an elementary combinatorial way, without any algebra. (In the construction of Lüneburg **[10]** the only algebraic tool is the existence of the group G.)

4. The projective space of dimension 3 and order 2. A *projective space of dimension* 3 *and order* 2 is an incidence structure $\Pi = (\tilde{\mathscr{P}}, \tilde{\mathscr{L}}, \tilde{I})$ with 15 points and 3 points on any line such that the following axioms hold:

(1) Any two distinct points of Π are incident with exactly one common line.

(2) (*Veblen-Young axiom*) Let l_1 and l_2 be two lines intersecting in a point. Let m_1 and m_2 be another two lines such that m_i intersects l_1 and l_2 in two distinct points $(1 \leqslant i \leqslant 2)$. Then m_1 and m_2 have a point in common. (Cf. Fig. 8.)

We shall now construct Π with the help of the projective plane of order 4, which we denote by $\mathbf{P} = (\mathscr{P}, \mathscr{L}, I)$, and one of its hyperovals $\mathscr{H}$. The set of all 3-lines of **P** is denoted by $\mathscr{T}$. Since there are exactly two 5-lines through any point of **P** outside $\mathscr{H}$, and since any two 5-lines intersect in a point outside $\mathscr{H}$, the following condition holds for any three 5-lines l, m, n: $l \cap m$, $l \cap n$, $m \cap n$ are three points of **P** outside $\mathscr{H}$ which are not collinear in **P**. We write

$$[l, m, n] = \{l \cap m, l \cap n, m \cap n\}$$

and call any such set a Δ-line. The set of all Δ-lines is denoted by $\mathscr{D}$:

$$\mathscr{D} = \{[l, m, n] \mid l, m, n \text{ are three distinct 5-lines of } \mathbf{P}\}.$$

In this section we shall show that the incidence structure

$$\Pi = (\mathscr{P} - \mathscr{H}, \mathscr{T} \cup \mathscr{D}, \in)$$

is the projective space of dimension 3 and order 2.

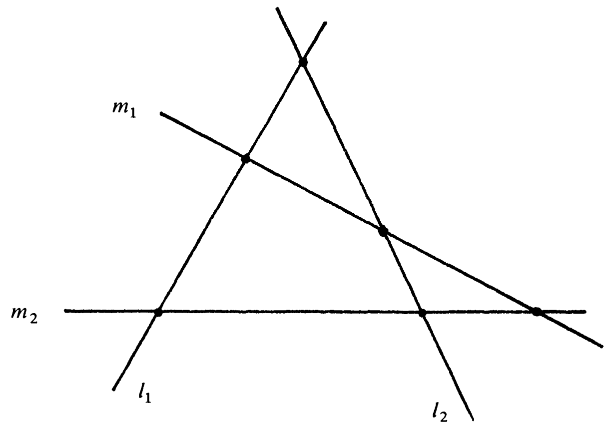


FIG. 8

By definition, Π has 15 points, and on any line of Π there are exactly 3 points. So, we have to verify axioms (1) and (2).

4.1 LEMMA. *Any two distinct points of* Π *are incident with a unique common line of* Π.

Proof. Denote by P_1 and P_2 two distinct points of Π. If P_1 and P_2 are connected in $\mathbf{P}$ by a 3-line m, then m is also the unique line of Π joining P_1 and P_2.

If P_1 and P_2 are not connected in $\mathbf{P}$ by a 3-line, then they are incident with a common 5-line l. Denote by l_i ($i = 1$ or 2) the 5-line distinct from l through P_i. Then $[l, l_1, l_2]$ is the Δ-line connecting P_1 and P_2. □

In order to prove that Π is a projective space, we must consider the planes of Π. For any point P of Π we define $\pi(P)$ as the set of points of Π (including P) which are connected with P by a 3-line. Since any point of Π is incident with three 3-lines, we have

$$|\pi(P)| = 1 + 3(3-1) = 7 \quad \text{for any point } P \text{ of } \Pi.$$

4.2 LEMMA. *For any point P of* Π, $\pi(P)$ *and the lines of* Π *contained in* $\pi(P)$ *form a projective plane of order* 2.

Proof. The proof is in two steps.

Step 1. Any two distinct points X_1, X_2 of $\pi(P)$ are on a common line of $\pi(P)$.

Denote by m_1, m_2, m_3 the 3-lines through P. Without loss in generality we have $X_1, X_2 \neq P$ and $X_1 \in m_1$, $X_2 \in m_2$. Since by 1.4 there can be no triangle of 3-lines in $\mathcal{P} - \mathcal{H}$, the line l of $\mathbf{P}$ through X_1 and X_2 is a 5-line.

Denote by l_i ($i = 1$ or 2) the 5-line different from l through X_i. The 5-line l_1 intersects m_3 in a point of Π which is different from P and $m_3 \cap l$, thus in the unique third point S of m_3 in Π. Similarly, l_2 passes through S. Thus, $[l, l_1, l_2] = \{X_1, X_2, S\}$. In other words, X_1 and X_2 are joined in Π by a line of $\pi(P)$.

Step 2. Any two lines of $\pi(P)$ intersect.

This follows by counting. Clearly, any point of $\pi(P)$ is incident with exactly 3 $(= (7-1)/(3-1))$ lines of $\pi(P)$. In other words, any line through a point X outside a line x intersects x. So, any two lines of $\pi(P)$ meet. With 7 points of which at most 3 are collinear, $\pi(P)$ must contain a quadrangle. Thus, $\pi(P)$ is a projective plane of order 2. □

4.3 LEMMA. *Through any three non-collinear points* X, Y, Z *of* Π *there is a unique plane of the form* $\pi(P)$.

Proof. Clearly, X, Y, Z are contained in at most one projective plane of order 2 in Π. So, we have only to show that there is a plane $\pi(P)$ through X, Y, Z.

If XY and XZ are 3-lines, then $X, Y, Z \in \pi(X)$. Suppose now that XY is a 3-line, while XZ is a 5-line. Let S be the third point in Π on XY. Since there are only two 5-lines through Z, either ZY or ZS is a 3-line. If (without loss in generality) $ZS \in \mathscr{T}$, then $X, Y, Z \in \pi(S)$.

Finally, suppose that XY and XZ are 5-lines. Since X, Y, Z are noncollinear in Π, it follows that YZ is a 3-line. Now the assertion follows as in the previous case. □

4.4 THEOREM. *Π is a projective space of dimension 3 and order 2.*

Proof. It remains only to show that Π satisfies the Veblen-Young axiom. Denote by l_1 and l_2 two lines of Π intersecting in a point X; suppose that m_1 and m_2 are two lines which do not contain X but intersect both l_1 and l_2. We must show that m_1 and m_2 meet.

By 4.3, l_1 and l_2 are contained in a common plane $\pi(P)$. Since m_1 and m_2 have two points in common with $\pi(P)$, m_1 and m_2 are lines of $\pi(P)$. By 4.2, m_1 and m_2 meet each other in a point of $\pi(P)$. □

REMARKS. 1. It follows from the proof of 4.4 that any plane of Π is of the form $\pi(P)$.

2. The projective space Π is usually denoted by $\Pi = PG(3,2)$. This incidence structure Π is unique (up to isomorphism). Indeed, the points and lines outside a fixed plane π_0 and Π form an "affine space"—which consists in our case just of the 8 vertices of a cube. Since this structure is unique and since one can adjoin the plane of infinity in only one way, π is unique as well.

The next Proposition tells us how we can recognize the Δ-lines in the graph K_6.

4.5 PROPOSITION. *Any Δ-line of Π is in 1-1 correspondence with a set of three disjoint factors of K_6 which are not contained in a common factorization.* □

The reader should convince himself that there exist such sets of factors in K_6.

4.6 COROLLARY. *The full automorphism group of Π has a subgroup isomorphic to $\mathscr{S}_6$.*

Proof. We know already that any permutation φ of $\mathscr{H}$ induces an automorphism of the incidence structure consisting of the points of Π and the 3-lines. By 4.5, φ also maps any Δ-line onto a Δ-line. □

REMARK. It is known that the full automorphism group of Π is isomorphic to $\mathscr{A}_8$, the alternating group on 8 symbols. (For a proof see, for instance, Conwell **[2]** or Pickert **[12]**.) It is obvious that Δ-lines and 3-lines are indistinguishable in Π even though it might not be evident in our model.

If we define the map π by

$$\pi : P \to \pi(P) \quad \text{and} \quad \pi : \pi(P) \to P,$$

then π is a bijection between the points and planes of Π. For this bijection π we have that

$$Q \in \pi(P) \Leftrightarrow P \in \pi(Q),$$

since both statements are equivalent to the fact that PQ is a 3-line. Such a bijection π is said to be a *polarity* of Π.

Note that $P \in \pi(P)$ for all points P; such a polarity is called *symplectic*. (Cf. for instance **[8]**.) If X, Y, Z are the points of Π on a 3-line l, then l is the common line of $\pi(X)$, $\pi(Y)$ and $\pi(Z)$. We call such a line *isotropic*. Now, let $[l, m, n]$ be a Δ-line. By the previous remark we know that $\pi([l, m, n])$ is not a 3-line. So, it is a Δ-line, $\pi([l, m, n]) = [l', m', n']$. We claim $\{l, m, n\} \cap \{l', m', n'\} = \varnothing$. Indeed, let P be the common point of l and m. Since $\pi(P)$ does not contain a line of the form $[l, m, x]$, we have $l \cap m \notin [l', m', n']$. Analogously, $l \cap n$, $m \cap n \notin [l', m', n']$. Hence l', m', n' are exactly those 5-lines which do not pass through the points $l \cap m$, $l \cap n$, $m \cap n$. In other words, $\{l, m, n\} \cap \{l', m', n'\} = \varnothing$.

To sum up these remarks,

4.7 THEOREM. (a) *The map* π *is a symplectic polarity of* Π.

(b) *The isotropic lines with respect to* π *are precisely the* 3-*lines*.

(c) *If* $[l, m, n]$ *is a* Δ-*line, then* $\pi([l, m, n]) = [l', m', n']$, *where this line is uniquely defined by* $\{l, m, n\} \cap \{l', m', n'\} = \varnothing$. □

5. Some geometric objects in Π. In this Section, we shall consider ovoids, reguli and spreads in Π. An *ovoid* of a 3-dimensional projective space of order n is a set $\mathcal{O}$ of $n^2 + 1$ points, no three of which are collinear, such that through any point P of $\mathcal{O}$ there is a plane (the *tangent plane*) with the property that it touches $\mathcal{O}$ only in P. Some ovoids of Π can be seen easily.

5.1 PROPOSITION. *The points of any* 5-*line of* $\mathbf{P}$ *form an ovoid of* Π.

Proof. No three points on a 5-line are collinear in Π. The tangent plane at the point P is $\pi(P)$, since this plane has only P in common with a 5-line through P. □

We call a set of lines of Π *skew*, if no pair of them intersect. Let Π_n be a 3-dimensional projective space of order n. A set $\mathcal{R}$ of $n + 1$ skew lines of Π_n is called a *regulus* if there is another set $\mathcal{R}'$ of $n + 1$ skew lines such that any line of $\mathcal{R}$ intersects any line of $\mathcal{R}'$. $\mathcal{R}'$ is called the *opposite* regulus.

It can be proved that in our projective space Π of order 2 any three skew lines form a regulus. (The reader is invited to prove this.) But—can you already "see" a regulus and its opposite regulus?

5.2 PROPOSITION. *If* $\mathcal{H} = \{A, B, C, D, E, F\}$ *is a hyperoval of* $\mathbf{P}$, *then* $\{AB, AC, BC\}$ *is a regulus of* Π *and* $\{DE, DF, EF\}$ *is its opposite regulus*.

Proof. We have already observed this fact in the first section, when we considered an affine plane of order 3 in $\mathbf{P}$. □

A *spread* of Π is a set $\mathcal{S}$ of lines such that any point of Π is incident with exactly one line of $\mathcal{S}$.

5.3 PROPOSITION. *The set of lines through a point* $A \in \mathcal{H}$ *is a spread of* Π. □

REMARK. It is not difficult to check that there are also spreads having three 3-lines and spreads having exactly one 3-line, but no other spreads.

Acknowledgements. The first version of this paper was written while the author was a visiting professor at the University of L'Aquila/Italy. He wants to thank Professor F. Eugeni and Professor L. Berardi for their kind hospitality. He wants to thank also Professor Chris Fisher for his careful reading of the manuscript, which improved both the mathematics and the English.

References

1. A. Bruen and R. Silverman, Arcs and blocking sets, London Mathematical Lecture Notes Series, 49 (1981) 52–60.

2. G. M. Conwell, The 3-space PG(3, 2) and its group, Ann. of Math., 11 (1910) 60–76.

3. H. S. M. Coxeter, Desargues configurations and their collineation groups, Math. Proc. Cambridge Philos. Soc., 78 (1975) 227–246.

4. W. L. Edge, Some implications of the geometry of the 21-point plane, Math. Z., 87 (1965) 348–362.

5. J. Hall, On identifying PG(3, 2) and the complete 3-design on seven points, Ann. Discrete Math., 7 (1980) 131–141.

6. F. Harary, Graph Theory, Addison-Wesley, Reading, MA, 1969.

7. J. W. P. Hirschfeld, Projective Geometries over Finite Fields, Clarendon Press, Oxford, 1979.

8. D. R. Hughes and F. C. Piper, Projective Planes, Springer, 1973.

9. W. Jonsson, On the groups M_{22}, M_{23}, M_{24} and the uniqueness of the associated Steiner systems, Math. Z., 125 (1972) 193–214.

10. H. Lüneburg, Über die Gruppen von Mathieu, J. Algebra, 10 (1968) 194–210.

11. T. G. Ostrom and F. A. Sherk, Finite projective planes with affine subplanes, Canad. Math. Bull., 7 (1964) 549–559.

12. G. Pickert, Von der Desargues-Konfiguration zum 5-dimensionalen projektiven Raum mit 63 Punkten, Math. Semesterber., 29 (1982) 51–67.

13. R.-H. Schulz, Geometrie über GF(4), Math.-Phys. Semesterber., 22 (1975) 158–184.

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

HOW MANY INTERSECTIONS CAN A HELICAL CURVE HAVE WITH THE UNIT SPHERE DURING ONE PERIOD?

THOMAS CARTER
Mathematics and CIS Department, Eastern Connecticut State University, Willimantic, CT 06226

Let us consider the following linear optimal control problem: Given a positive real number t_f, let $\mathscr{U}$ denote the set of all Lebesgue measurable functions that map the closed interval $T = [0, t_f]$ into the closed unit ball U in $\mathbb{R}^3$. The elements of $\mathscr{U}$ will be referred to as *control functions*. We seek $u \in \mathscr{U}$ to minimize the functional

$$J[u] = \int_0^{t_f} |u(t)|\, dt,$$

where $|\ |$ denotes the Euclidean norm or magnitude, subject to the differential equations

$$\dot{x}(t) = v(t),$$
$$\dot{v}(t) = Ax(t) + Bv(t) + bu(t),$$

which hold a.e. on T, where b is a positive real number, the dot denotes differentiation with respect to time $t \in T$, and the boundary conditions

$$x(0) = x_0, \qquad x(t_f) = x_f,$$
$$v(0) = v_0, \qquad v(t_f) = v_f,$$

where x_0, v_0, x_f, and v_f are specified points in $\mathbb{R}^3$. The matrices A and B are of the form

$$A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & (k^2-1)\Omega & 0 \\ 0 & 0 & -\Omega^2 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & k\Omega & 0 \\ -k\Omega & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where Ω and k are real numbers with $\Omega > 0$ and $k \neq \pm 1$.

Since T and U are compact, the problem defined above is a fixed time bounded control problem. It is well known in the literature of optimal control that if a function $u \in \mathscr{U}$ exists, such that the differential equations and boundary conditions are satisfied, then there also exists a control function in $\mathscr{U}$ such that these conditions are satisfied and J is minimized. In this case the

well-known principle of Pontryagin [1] provides both necessary and sufficient conditions for a minimum because the integrand of the cost function J is independent of x and v [2, Sect. 5.2]. For this reason the triple (x, v, u) is a minimizing solution of the preceding optimal control problem if and only if the differential equations and boundary conditions are satisfied and the Hamiltonian function

$$H(x(t), v(t), p(t), q(t), \cdot): U \to \mathbb{R}$$

defined by

$$H(x(t), v(t), p(t), q(t), r) = l|r| + p(t)^T v(t) + q(t)^T Ax(t) + q(t)^T Bv(t) + bq(t)^T r$$

is minimized by $u(t)$ a.e. on T where $l \geqslant 0$ and the functions p and q which map T into $\mathbb{R}^3$ are absolutely continuous solutions of the adjoint differential equations

$$\dot{p}(t) = -H_x(x(t), v(t), p(t), q(t), u(t)) = -A^T q(t),$$
$$\dot{q}(t) = -H_v(x(t), v(t), p(t), q(t), u(t)) = -B^T q(t) - p(t),$$

where the subscripts denote that the gradient is taken with respect to these vectors and the superscript ^{T} is used to indicate that a matrix is transposed.

For those boundary conditions in which $l = 0$ the minimizing solutions are called *abnormal*, otherwise the minimizing solutions are called *normal*. For minimizing solutions that are normal l may be any positive number. For convenience we set $l = b$.

If the Hamiltonian function has a unique minimum on U a.e. on a Lebesgue measurable set $S \subseteq T$, then the minimizing solution is called *nonsingular* on S. If it has non-unique minima a.e. on a measurable set S, then the minimizing solutions are called *singular* on S.

An examination of the Hamiltonian function shows that a normal minimizing solution is singular on a set $S \subseteq T$ of positive measure if and only if $|q(t)| = 1$ a.e. on S. It is nonsingular on a set S of positive measure if and only if $|q(t)| \neq 1$ a.e. on S, and in this case the minimizing control function is defined a.e. on S by

$$\begin{aligned} u(t) &= 0, && |q(t)| < 1 \\ u(t) &= -q(t)/|q(t)|, && |q(t)| > 1. \end{aligned}$$

This expression shows that a normal nonsingular minimizing control function has values that are either on the unit sphere or are zero. Moreover the function q determines entirely which of these situations occurs or whether or not a minimizing solution is singular. If we eliminate p in the adjoint differential equations we see that q satisfies the differential equation

$$\ddot{q}(t) + B^T \dot{q}(t) - A^T q(t) = 0.$$

This differential equation can be readily solved. Using subscripts to denote the components of q and abusing the notation for q we obtain

$$\begin{aligned} q_1(\theta) &= k\rho \sin\theta + c_1\theta + c_2, \\ q_2(\theta) &= \rho\cos\theta + \frac{kc_1}{k^2 - 1}, \\ q_3(\theta) &= \alpha\sin\theta + \beta\cos\theta, \end{aligned}$$

where $\theta = \Omega t + \psi$ and ψ, ρ, c_1, c_2, α, and β are arbitrary real constants of integration with $\rho \geqslant 0$. It can be shown from the form of q that a normal minimizing solution is either singular on T or nonsingular on T (i.e., there are no singular arcs joining nonsingular arcs). If $c_1 \neq 0$, a minimizing solution is nonsingular on T.

We note that the *switching function* s: $T \to \mathbb{R}$ defined by $s(t) = |q(t)| - 1$ determines a.e. on T whether a normal nonsingular minimizing control function u has values on the unit sphere or at

the origin in $\mathbb{R}^3$ (i.e., $u(t) = 0$ if $s(t) < 0$ and $u(t) = -q(t)/|q(t)|$ if $s(t) > 0$ a.e. on T). A point $t_s \in T$ at which the switching function s changes sign will be called a *switching time*. Since q is continuous we see that $s(t_s) = 0$ at any switching time $t_s \in T$. We observe that a normal minimizing solution is singular if and only if the switching function is identically zero on T.

We now raise the following conjecture: *Any normal nonsingular minimizing control function has at most six switching times if* $t_f \leqslant 2\pi/\Omega$.

This conjecture about solutions of the optimal control problem is easily found to be true if the following slightly more general statement is true: *The switching function associated with any normal minimizing solution is either identically zero or has at most six zeros if* $t_f \leqslant 2\pi/\Omega$. It is this statement that we would like to establish. It is easily seen to be equivalent to the following conjecture in geometry: *If, for* $\psi \leqslant \theta \leqslant \psi + 2\pi$, *the curve* $q(\theta)$ *defined above is not identically on the unit sphere in* $\mathbb{R}^3$, *then it can intersect the unit sphere on at most six points.*

This conjecture is clearly true for certain degenerate cases. It is not difficult to see that if $c_1 = 0$ the curve defined by $q(\theta)$ is a segment of an ellipse (possibly degenerate) in real three space and can therefore either intersect the unit sphere on at most four points during one period ($\psi \leqslant \theta \leqslant \psi + 2\pi$) or else degenerates to a circular segment or a fixed point which is contained entirely on the unit sphere in $\mathbb{R}^3$. In the former case there are at most four switching times in one period, and in the latter case the minimizing solutions are singular. Similarly if $\rho = 0$ the curve is sinusoidal and in this case the equation $|q(\theta)| = 1$ can be shown to have at most six roots during one period. In this case, it is not difficult to see that the maximum of six roots in one period can occur. This happens for example, if $\psi = -\pi$, $c_1 = 1/\pi$, $\alpha = 1$, and $k = c_2 = \beta = 0$. If α and β are both zero, the curve degenerates to a straight line and at most two roots are possible.

Since the conjecture is clearly true in the preceding cases, we shall assume c_1 and ρ are nonzero and restate it as the following problem in geometry: *Given real numbers* $k \neq \pm 1$, $\rho > 0$, $c_1 \neq 0$, *and* c_2, α, β, *and* ψ *arbitrary, show that the curve* $q(\theta) = (q_1(\theta), q_2(\theta), q_3(\theta))$ *defined by the equations*:

$$q_1(\theta) = k\rho \sin\theta + c_1\theta + c_2,$$

$$q_2(\theta) = \rho\cos\theta + \frac{kc_1}{k^2 - 1},$$

$$q_3(\theta) = \alpha\sin\theta + \beta\cos\theta,$$

can intersect the unit sphere in $\mathbb{R}^3$ *on at most six points for* $\psi \leqslant \theta \leqslant \psi + 2\pi$.

The following special cases are of particular interest.

1. $k = \beta = 0$. This is the problem of determining the number of intersections of a standard elliptical helix with the unit sphere. The conjecture can be shown to be true in this case.
2. $\alpha = \beta = 0$. This is the planar problem of determining the number of intersections of a cycloidal curve with the unit circle. In real two space the curve defined by $(q_1(\theta)/k, q_2(\theta))$ is a type of cycloid if $k \neq 0$. It is a prolate cycloid if $\rho > |c_1/k|$ and a curtate cycloid if $\rho < |c_1/k|$.
3. $k = 2$. In this case $x(t)$ and $v(t)$ represent the position and velocity of a spacecraft measured with respect to a coordinate system which is fixed in a satellite which moves in a circular orbit about a planet. This model was originally discovered by Wheelon [3] and also independently by Clohessy and Wiltshire [4]. In this model the thrust acceleration vector is given by $bu(t)$, where b represents the maximum magnitude of the thrust acceleration. The bounded control function $u \in \mathscr{U}$ therefore represents a normalized thrust acceleration function. The spacecraft is assumed to have a constant point mass, and an inverse square law gravitational force function has been linearized with respect to a point which represents the location of the satellite in circular orbit. The positive number Ω represents the angular speed of the satellite. We are assuming constant exhaust velocity for the spacecraft so that the fuel consumption rate is proportional to the magnitude of the thrust, hence $J[u]$ is proportional to the total fuel consumed during the flight

time t_f, and a control function that minimizes $J[u]$ subject to the other restrictions is a minimum fuel control function.

The last case falls in the general field of optimal space trajectories, an area pioneered by Derek Lawden [**5**]. The function q, which he called the *primer vector*, appears in the case of a spacecraft coasting in Keplerian orbit [**5**, equations 5.45–5.47]. A more recent description of optimal space trajectories can be found in the work of Marec [**6**] which includes, in particular, the minimum fuel trajectories of a spacecraft near a satellite in which the linearization is performed through the use of orbital elements. The same basic results can be found using this model as are found with the Clohessy-Wiltshire equations, but the approach is more general.

The problem is discussed in [**7**] with emphasis on the planar case in which $\alpha = \beta = 0$. An example is presented of a planar primer vector curve q which intersects the unit circle at six distinct points during one period of revolution of the satellite. It is claimed that q cannot intersect the unit sphere on more than six points during one period of revolution of the satellite unless it coincides identically with the unit sphere. In this case the optimal trajectory is singular and is discussed at length in [**8**]. On the basis of this claim it is asserted that normal nonsingular fuel-optimal maneuvers of a spacecraft near a satellite in circular orbit about a planet consist of at most seven intervals of either full thrust or coast during one period of revolution of the satellite.

References

1. L. S. Pontryagin, V. G. Boltyanski, R. V. Gamkrelidze, and E. F. Mishchenko, The Mathematical Theory of Optimal Processes, Wiley, New York, 1962.

2. E. B. Lee and L. Markus, Foundations of Optimal Control Theory, Wiley, New York, 1967.

3. A. D. Wheelon, Midcourse and Terminal Guidance, Space Technology, Wiley, New York, 1959.

4. W. H. Clohessy and R. S. Wiltshire, Terminal Guidance System for Satellite Rendezvous, J. Aerospace Sciences, Sept. 1960, pp. 653–658, 674.

5. D. F. Lawden, Optimal Trajectories for Space Navigation, Butterworths, London, 1963.

6. J. P. Marec, Optimal Space Trajectories, Elsevier, New York, 1979.

7. T. E. Carter, Fuel-Optimal Maneuvers of a Spacecraft Relative to a Point in Circular Orbit, AIAA J. Guidance, Control, and Dynamics, Nov.–Dec. 1984.

8. T. E. Carter, Singular Fuel-Optimal Space Trajectories Based on a Linearization about a Point in Circular Orbit, J. Optim. Theory Appl. (accepted for publication).

NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.

For instructions about submitting Notes for publication in this department see the inside front cover.

ON JENSEN'S FORMULA AND $\int_0^{2\pi} \log|1 - e^{i\theta}|\,d\theta$

ROBERT M. YOUNG

Department of Mathematics, Oberlin College, Oberlin, OH 44074

One of the fundamental theorems of analysis, Jensen's formula establishes an intimate connection between the values of an analytic function on a circle and the distribution of its zeros inside the circle. Suppose that $f(z)$ is analytic for $|z| < R$ and that $f(0) \neq 0$. If $0 < r < R$ and if $z_1, z_2, \ldots, z_n$ are the zeros of $f(z)$ in $|z| \leqslant r$, listed according to their multiplicities, then

$$\frac{1}{2\pi}\int_0^{2\pi} \log|f(re^{i\theta})|\,d\theta = \log|f(0)| + \sum_{k=1}^{n} \log\left(\frac{r}{|z_k|}\right).$$

The standard proof of Jensen's formula ultimately involves showing that

$$\int_0^{2\pi} \log|1 - e^{i\theta}|\, d\theta = 0,$$

which is then established by means of the residue theorem (see, e.g., [1], [2], [3], and [4]). But residues are unnecessary—the integral is in fact completely elementary, as the following calculations show.

Since $|1 - e^{i\theta}| = 2\sin\frac{\theta}{2}$, we need only verify that

$$\int_0^{\pi} \log \sin x\, dx = -\pi \log 2.$$

Call the left side I and write $\sin x = 2\sin\frac{x}{2}\cos\frac{x}{2}$. Then

$$I = \pi\log 2 + \int_0^{\pi} \log\sin\frac{x}{2}\, dx + \int_0^{\pi}\log\cos\frac{x}{2}\, dx.$$

Replacing $\frac{x}{2}$ by t in the first integral and by $\frac{\pi}{2} - t$ in the second, we find

$$\begin{aligned} I &= \pi\log 2 + 4\int_0^{\pi/2}\log\sin t\, dt \\ &= \pi\log 2 + 2I, \end{aligned}$$

and the result follows.

References

1. L. V. Ahlfors, Complex Analysis, 3rd ed., McGraw-Hill, 1979.
2. J. B. Conway, Functions of One Complex Variable, 2nd ed., Springer-Verlag, 1978.
3. N. Levinson and R. M. Redheffer, Complex Variables, Holden-Day, 1970.
4. W. Rudin, Real and Complex Analysis, 2nd ed., McGraw-Hill, 1973.

ANTISOCIAL SUBCOVERS OF SELF-CENTERED COVERINGS

Steven G. Krantz* and T. D. Parsons
Department of Mathematics, Pennsylvania State University, University Park, PA 16802

1. Introduction. It is a remarkable and relatively recent discovery that fundamental results in measure theory are logically equivalent to certain covering properties of balls in Euclidean space. For instance, consider the Lebesgue Differentiation Theorem, which says that if f is a Lebesgue integrable function on $\mathbb{R}^n$, m is Lebesgue measure on $\mathbb{R}^n$, and $B(x,r)$ is the ball $\{t \in \mathbb{R}^n: |x - t| < r\}$, then

$$\lim_{r\to 0^+} \frac{1}{m(B(x,r))}\int_{B(x,r)} f(t)\, dt = f(x)$$

for almost every $x \in \mathbb{R}^n$. This result is logically equivalent to the following covering theorem:

Theorem A. *Let K be a compact subset of $\mathbb{R}^n$ and let $\mathscr{C} = \{B(x_i, r_i): i \in I\}$ be a collection of balls which cover K. Then there is a finite subcollection $\{B(x_j, r_j): j \in J\} \subseteq C$, whose members are pairwise disjoint, such that $K \subset \bigcup_{j\in J} B(x_j, 3r_j)$.* (See [3], [1] for details on these results.)

Thus covering theorems like Theorem A play a central role in modern real and harmonic analysis. In this note we prove a covering theorem of a slightly different nature and then derive a surprising theorem about partitions of covers. Theorem A holds in any metric space. See [2] and references therein.

*Research supported in part by a grant from the National Science Foundation.

Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

T:	Textbook	13-18:	Grade Level
S:	Supplementary Reading	1-4 :	Time in Semesters
P:	Professional Reading	** :	Special Emphasis
L:	Undergraduate Library	?? :	Questionable

General, P, L. New Directions in Two-Year College Mathematics. Ed: Donald J. Albers, Stephen B. Rodi, Ann E. Watkins. Springer-Verlag, 1985, xx + 491 pp, $24. [ISBN: 0-387-96145-3] Proceedings of a July 1984 conference held at Menlo College, California on issues of curriculum, pedagogy, and faculty development in two-year colleges. Groups of papers are followed by edited highlights of the discussion they generated. A timely, thought-provoking volume. LAS

General, T(13-14: 1, 2), S, L. Mathematical Applications for Management, Life, and Social Sciences, Second Edition. Ronald J. Harshbarger, James J. Reynolds. DC Heath, 1985, xii + 755 pp, $29.95. [ISBN: 0-669-07337-7] This new edition has a new presentation of the simplex method, reorganized material on the mathematics of finance, a discussion of Markov chains, and an increased coverage of calculus. Presupposes two years of high school algebra. (First Edition, TR, August-September 1981) FLW

General, P. The Influence of Computers and Informatics on Mathematics and Its Teaching. International Commission on Mathematical Instruction. IREM (US Distr: MAA), 1985, 321 pp, (P). Complete collection of papers presented at a March 1985 conference held in Strasbourg, France under the sponsorship of ICMI to examine from an international perspective the effects of computing on mathematics research and mathematics teaching. Includes papers on symbolic algebra, geometry, discrete mathematics and computer graphics. LAS

General, S(15-18), L. The Fourth Dimension and Non-Euclidean Geometry in Modern Art. Linda Dalrymple Henderson. Princeton U Pr, 1983, xxiii + 453 pp, $60; $22.50 (P). [ISBN: 0-691-04008-7; 0-691-10142-6] A remarkable history of the influence of mathematics on art of the past century--especially painting--by an art historian knowledgeable about mathematics. SS

Elementary, T(13: 2). Intermediate Algebra. Andrew Demetropoulos, Kenneth C. Wolff. Macmillan, 1985, xiii + 571 pp. [ISBN: 0-02-328530-3] Written for college students. Topics are introduced with "everyday" applications; provides word problems and calculator exercises. Instructor's Manual and Student Solutions Manual available. JNC

Mathematics Appreciation, S(13-15), L*. The Beauty of Doing Mathematics: Three Public Dialogues. Serge Lang. Springer-Verlag, 1985, x + 127 pp, $19.80 (P). [ISBN: 0-387-96149-6] English translation of transcripts from three highly interactive public lectures delivered annually by Lang at the Palais de la Découverte (Science Museum) in Paris. A remarkably intelligent lay audience follows Lang deep into the territory of prime numbers, diophantine equations, and geometric topology, asking questions and suggesting conjectures at every turn. LAS

Precalculus. Technical Mathematics with Applications. C.E. Goodson, S.L. Miertschin. Wiley, 1983, xii + 1002 pp, $25.95. [ISBN: 0-471-08244-9] Usual precalculus topics for college-level students in technical, pre-engineering and career-oriented programs; numerous traditional word problems plus optional sections devoted to specific applications; chapters begin with lists of competencies and diagnostic self-tests; appendix on use of calculator and computer. JNC

Precalculus, T*(13). Algebra and Trigonometry. Roland E. Larson, Robert P. Hostetler. DC Heath, 1985, xvi + 648 pp, $27.95. [ISBN: 0-669-05644-8] The standard topics of precalculus along with an introduction to systems of equations, matrices and determinants, sequences, series, and probability. This volume does not include analytic geometry. Over 700 examples, 4000 exercises and 900 figures. Emphasizes the algebra of calculus and includes calculator exercises. CEC

Precalculus, T*(13: 1). Trigonometry and its Applications. Christian R. Hirsch, Harold L. Schoen. McGraw-Hill, 1985, xiii + 528 pp, $28.95. [ISBN: 0-07-029059-8] If you offer a course in trigonometry, as opposed to algebra and trigonometry, this book offers many helpful pedagogical features. Calculator use is encouraged. Each chapter has an optional section on BASIC programs. The last chapter is a light treatment of sequences, series and hyperbolic functions. Appendix A is

on BASIC programming. JK

Precalculus, T(13: 1). College Algebra, Fourth Edition. Margaret L. Lial, Charles D. Miller. Scott Foresman, 1985, 501 pp, $25.95. [ISBN: 0-673-18024-7] New to this edition (First Edition, TR, October 1973; Second Edition, TR, August-September 1978; Third Edition, TR, August-September 1981) are sections on probability theory and business applications; supplements include an Instructor's Guide containing a diagnostic pretest covering the first three chapters and audiotapes. JNC

Finite Mathematics, T*(13: 1). Applied Finite Mathematics. George J. Kertz. West, 1985, xiii + 530 pp [ISBN: 0-314-85255-7]; Instructor's Manual with Test Bank to Accompany, v + 261 pp, (P). [ISBN: 0-314-87238-8] Matrices, linear programming, probability, finance, logic; some statistics; linear, quadratic, exponential, logarithmic functions. Brief introduction to graph theory. Attractive and very readable, with an abundant supply of examples and exercises which make the student think. Instructor's Manual consists of solutions and sample tests. DFA

Finite Mathematics, T(13-14: 1). Finite Mathematics, Third Edition. Margaret L. Lial, Charles D. Miller. Scott Foresman, 1985, 518 pp, $26.95. [ISBN: 0-673-18023-9] New section on duality in the chapters on the simplex method; new section on sequences in the chapter on mathematics of finance (to allow for coverage of annuities and amortization); new chapter on digraphs and networks. (Second Edition, TR, August-September 1982.) LCL

Education, P, L. Indicators of Precollege Education in Science and Mathematics: A Preliminary Review. Ed: Senta A. Raizen, Lyle V. Jones. National Academy Pr, 1985, ix + 200 pp, (P). [ISBN: 0-309-03536-8] A report based on currently available data of assessment measures--indicators--that can provide a basis for monitoring current reforms in science and mathematics education. Data on teachers, curriculum, enrollments, achievement tests are examined both at national and state levels. A useful compendium of existing information. LAS

Education, P, L.** The Secondary School Mathematics Curriculum, 1985 Yearbook. Christian R. Hirsch, Marilyn J. Zweng. NCTM, 1985, vi + 250 pp, $14.50. [ISBN: 0-87353-217-1] A timely collection of articles on various reforms under way or under study in secondary school mathematics, led off by a call for "revolution" by Zal Usiskin. 21 articles explore new curricular directions and innovative programs, including discrete and computer mathematics, integrated programs, and special curricula. LAS

Education, P, L. Math! Encounters with High School Students. Serge Lang. Springer-Verlag, 1985, xii + 138 pp, $19.95 (P). [ISBN: 0-387-96129-1] Transcript of Socratic dialogues with high-ability high school students (in Canada and France) on problems of volume, area, and length; Pythagorean triples; and infinities. Extensive dialogue, conjectures, bad guesses, and corrections make this volume more of interest as a case study in mathematical instruction than as an exposition of the topics presented. LAS

History, S, P*, L*. Ramanujan's Notebooks, Part I. Bruce C. Berndt. Springer-Verlag, 1985, x + 357 pp, $54. [ISBN: 0-387-96110-0] The first of three volumes devoted to the editing of Ramanujan's notebooks. Proof of each of Ramanujan's theorems is given, except for known results for which references are given. A tribute to Ramanujan's genius and the author's perseverence. CEC

History, S(15-17), P, L.** Enigmas of Chance: An Autobiography. Mark Kac. Harper & Row, 1985, xxvii + 163 pp, $18.95. [ISBN: 0-06-015433-0] A posthumously published autobiography, one in the series of Sloan Foundation "lives in science." Kac's account is reserved and concise, interleaving his biography (fron Lwów to Cornell to Rockefeller) with his life-long search for the meaning and applications of independence--in probability, in number theory, in physics, in communication theory. Although rich in anecdote and amply enhanced by exposition of mathematics, Kac's evocation of his life in science is nonetheless two-dimensional, rarely revealing the human emotions that surely accompanied his distinguished career. LAS

Graph Theory, T(17: 1), S, P*. Interval Orders and Interval Graphs: A Study of Partially Ordered Sets. Peter C. Fishburn. Ser. in Discrete Math. Wiley, 1985, xi + 215 pp, $33.50. [ISBN: 0-471-81284-6] A unified, self-contained introduction to interval orders, interval graphs and related concepts such as semiorders, comparability graphs, and indifference graphs. An outgrowth of the authors research. Includes an excellent list of references but no exercises. CEC

Combinatorics, T, S*, P, L*. The Equivalence of Some Combinatorial Matching Theorems. Philip F. Reichmeider. Polygonal, 1984, 123 pp, $15.50. [ISBN: 0-936428-09-0] An exploration of a class of combinatorial matching theorems with special emphasis on the classical theorems of Hall, König, Dilworth, Menger, Ford and Fulkerson. A number of proofs of each together with a discussion of interrelationships. LCL

Combinatorics, P. Orders: Description and Roles. Ed: Maurice Pouzet, Denis Richard. Annals of Discrete Math., V. 23. Elsevier Science, 1984, xxvii + 548 pp, $69 (P). [ISBN: 0-444-87601-4] The 27 papers in this volume survey various aspects of the theory of order. They represent the proceedings of the Conference on Ordered Sets and Their Applications which was held at l'Arbresle in June, 1982. CEC

Number Theory, T*(14: 1), S, P, L*. Elementary Number Theory and Its Applications. Kenneth H. Rosen. Addison-Wesley, 1984, xii + 452 pp, $29.95. [ISBN: 0-201-06561-4] An introductory course that includes material on recursion, algorithms and complexity, computer arithmetic with large

integers, binary and hex representations, primality testing, hashing functions and public-key cryptography along with standard topics in number theory. Includes lots of exercises and computer projects. A well-written text. CEC

Number Theory, T*(14: 1), S, P*, L*. A Concise Introduction to the Theory of Numbers. Alan Baker. Cambridge U Pr, 1984, xiii + 95 pp, $29.95; $9.95 (P). [ISBN: 0-521-24383-1; 0-521-28654-9] Although concise, this book is thorough and self-contained. It is amazing that the author can accomplish so much in 95 pages without losing readability. Includes excellent exercises. The material in this book makes up a superb introduction to number theory. It is a gem! CEC

Linear Algebra, T(14-15: 1), S, L*. Applied Matrix Models: A Second Course in Linear Algebra with Computer Applications. Andy R. Magid. Wiley, 1985, x + 240 pp, $32.95. [ISBN: 0-471-88865-6] A textbook for a second course in linear algebra emphasizing applications and the use of standard libraries of numerical software (LINPACK and EISPACK). An interesting blend of theory and practice. AO

Group Theory, T(17: 1), S, P. Representations of Compact Lie Groups. Theodor Bröcker, Tammo tom Dieck. Grad. Texts. in Math., V. 98. Springer-Verlag, 1985, x + 313 pp, $39. [ISBN: 0-387-13678-9] Introduction to the representation theory of compact Lie groups using classical approach of Weyl, geometric and analytic methods stressed, Lie algebras secondary. Emphasis on the character ring, representative functions and the Peter-Weyl theorem, maximal tori, root systems, weights, the Weyl character formula. Nice exercises, discussion of the classical examples. RM

Group Theory, T*(16-17: 1), S*, P*, L*. Finite Reflection Groups, Second Edition. L.C. Grove, C.T. Benson. Grad. Texts in Math., V. 99. Springer-Verlag, 1985, x + 133 pp, $24. [ISBN: 0-387-96082-1] Additions to the First Edition include a chapter on classical invariant theory of finite reflection groups and an appendix listing the Schoenflies and international notational schemes for crystallographic point groups. For upperclassmen and graduate students. Approach is algebraic with a geometric flavor to achieve "...a middle ground between Coxeter and Bourbaki." JK

Algebra, P. Tables of Dominant Weight Multiplicities for Representations of Simple Lie Algebras. M.R. Bremner, R.V. Moody, J. Patera. Pure & Appl. Math., V. 90. Dekker, 1985, v + 340 pp, $78. [ISBN: 0-8247-7270-9] After a brief textual explanation and summary, the substance of the book is computer-generated tables for Lie algebra representations, mostly positive roots tables and multiplicity tables. References. JS

Algebra, T*(17: 2), S, P, L*. Basic Algebra I, Second Edition. Nathan Jacobson. WH Freeman, 1985, xviii + 499 pp, $35.95. [ISBN: 0-7167-1480-9] Two important changes occur in the chapter on Galois theory: a completely new section on finite fields and a different proof of the formula for the number of monic irreducible polynomials of degree n in a finite field. Also gives a new proof of the basic elimination theorem. An appendix which lists topics for independent study has been added, as have additional exercises. A solid introduction to algebra for the advanced undergraduate or beginning graduate student (First Edition, TR, January 1975). CEC

Calculus, T(13: 1). Applied Calculus. Alan M. Baum, Stephen J. Milles, Henry J. Schultz. Wiley, 1985, xii + 364 pp, $25.95. [ISBN: 0-471-80306-5] For business and biology students; emphasis is on skills and techniques. Review of algebra, differentiation and integration, exponential and logarithmic functions, functions of several variables. Many examples, drill problems, applied problems. Useful for classes with different levels of preparation. DFA

Calculus, T(13-14: 3). Calculus with Analytic Geometry. Richard A. Silverman. Prentice-Hall, 1985, xvi + 1055 pp. [ISBN: 0-13-111634-7] Much optional material (for "versatility") and a "patient explanation" style to enhance accessibility yields another massive text. Includes applied problems in economics, astrophysics, biomedical science and information theory; introduces limits and continuity simultaneously; early presentation of L'Hôpital's rule. JNC

Calculus, T*(13-14: 3). Calculus with Analytic Geometry, Second Edition. John B. Fraleigh. Addison-Wesley, 1985, xvi + 1017 pp. [ISBN: 0-201-12010-0] New two-color format and numerous additional figures greatly enhance the appeal of this edition (First Edition, TR, August-September 1980). Includes many more exercises and examples. JNC

Real Analysis, T(17-18: 1), S, P, L. The Geometry of Fractal Sets. K.J. Falconer. Tracts in Math., V. 85. Cambridge U Pr, 1985, xiv + 162 pp, $32.50. [ISBN: 0-521-25694-1] This 150-page book is filled with the geometric jewels of fractional and integral Hausdorff dimension. It contains a much-needed unified notation and includes many recent results with simplified proofs. The theory is classically and rigorously presented with applications only alluded to in the introduction. Each chapter contains a short and important problem set. This is a lovely introduction to the mathematics of fractal sets for the pure mathematician. PH

Real Analysis, T(16-17: 2), S, L. An Introduction to Analysis and Integration Theory. Esther R. Phillips. Dover, 1984, xxviii + 452 pp, $10.95 (P). [ISBN: 0-486-64747-1] Reprint of a 1971 volume published by Intext (TR, January 1972). In this edition a new historical introduction has been added. The unifying theme of the text is the completion of a space: the completion of the rational numbers and then generalized metric spaces and normed linear spaces dominates Part I. In Part II the Lebesgue integrable functions are constructed by completing a space of step functions. Part III contains an introduction to Banach and Hilbert spaces, and Part IV introduces the Daniell integral and measure spaces. Inexpensive, clearly written, with numerous exercises. MU

Complex Analysis, T(16-17: 1), S, P, L. Complex Analysis with Applications. Richard A. Silverman. Dover, 1984, x + 274 pp, $6.50 (P). [ISBN: 0-486-64762-5] A quite complete and careful introduction to the theory, through Schwarz-Christoffel formulas and harmonic functions, up to but not including Riemann mapping theorem, normal families. Theorems (e.g., Cauchy-Goursat) are proved in general forms. Integration treated early, standard transcendental and algebraic function examples later. Attractive treatment of physical applications. Helpful summaries and remarks end chapters. Numerous, varied exercises. Reprint of the 1975 Prentice-Hall edition (TR, November 1974; Extended Review, January 1976). PZ

Complex Analysis, P. Topics in Several Complex Variables. Ed: E. Ramírez de Arellano, D. Sundararaman. Research Notes in Math., V. 112. Pitman, 1985, 189 pp, $18.95 (P). [ISBN: 0-273-08656-1] Proceedings of the Workshop on Several Complex Variables held in the Centro Vacacional "La Trinidad," Santa Cruz, Mexico in August 1983. BH

Numerical Analysis, S(16), P. Introduction to Numerical Methods for Parallel Computers. U. Schendel. Transl: B.W. Conolly. Ser. in Math. & Its Applic. Halsted Pr, 1984, 151 pp, $21.95. [ISBN: 0-470-20091-X] A brief introduction for scientists, engineers and mathematicians. Discusses models of parallel computation, construction of algorithms, and special methods for linear systems, eigenvalue problems and nonlinear equations. RWN

Numerical Analysis, T(14: 1). Elementary Numerical Analysis. Kendall Atkinson. Wiley, 1985, xii + 416 pp, $31.95. [ISBN: 0-471-89733-7] A textbook for a lower-division introductory course in numerical analysis. Includes the standard topics: errors, rootfinding, interpolation, approximation, integration, differentiation, linear systems of equations, and differential equations. FORTRAN 77 programs for many of the algorithms are included. AO

Numerical Analysis, P. Elliptic Problem Solvers II. Ed: Garrett Birkhoff, Arthur Schoenstadt. Academic Pr, 1984, xiii + 573 pp, $39. [ISBN: 0-12-100560-7] The proceedings of a conference held at the Naval Postgraduate School in January 1983. The papers in this volume discuss various aspects of the "state-of-the-art" in the numerical solution of elliptic boundary value problems. AO

Functional Analysis, S(18), P. Lecture Notes in Mathematics-1113: Matrix Methods in Analysis. Piotr Antosik, Charles Swartz. Springer-Verlag, 1985, iv + 114 pp, $9.80 (P). [ISBN: 0-387-15185-0] These notes present a result which is simpler than the Antosik-Mikusinski Diagonal Theorem and which can be used to treat a wide variety of topics in functional analysis and measure theory, e.g., the Uniform Boundedness Principle, the Banach-Steinhaus Theorem, and much more. Reader-friendly, not presuming a great deal of specialization. MU

Analysis. Lecture Notes in Mathematics-1062: Harmonic Maps Between Surfces. Jürgen Jost. Springer-Verlag, 1984, x + 133 pp, $8 (P). [ISBN: 0-387-13339-9] Topics include existence and uniqueness theorems for harmonic maps between surfaces, existence theorems for harmonic diffeomorphisms, and $C^{1,\alpha}$ and $C^{2,\alpha}$ estimates for harmonic maps. Also considers harmonic maps in general. BH

Analysis, P. Polyharmonic Functions. Nachman Aronszajn, Thomas M. Creese, Leonard J. Lipkin. Oxford U Pr, 1983, x + 265 pp, $59. [ISBN: 0-19-853906-1] A function u defined on a domain D in R^n is polyharmonic on D if its sequence of Laplacians $\{\Delta^p u\}_{p\geq 1}^{\infty} = 0$ satisfies a certain growth condition on compact subsets of D. This monograph develops the theory and applications of these functions, including analytic continuation, singularities, local Laplacian order and type, and the Almansi expansion. Includes appendix on holomorphic capacity and an extensive bibliography. BH

Analysis, T(15-16: 1-3), S, P, L. Unified Integration. E.J. McShane. Pure & Appl. Math., V. 107. Academic Pr, 1983, xiii + 607 pp, $55. [ISBN: 0-12-486260-8] Studies the "gauge integral," which is defined using the notions of gauge (a certain interval-valued function) and gauge-fine partition. The Lebesgue integral turns out to be equivalent, and the Riemann integral to be a special case; the latter need not be abandoned in favor of the former. Contains material and exercises for several courses in applications and extensions of the theory: differential equations, multivariable calculus, line integrals and surface area, orthogonal expansions, measure theory. Lively, inviting, discursive style. PZ

Analysis, T(16-17: 1), L. Introduction to Applicable Mathematics, Part II: Advanced Analysis. Fred A. Hinchey. Halsted Pr, 1984, vii + 441 pp, $29.95. For science and engineering students; second in a three-volume series. Conformal mapping, integral transforms, integral and asymptotic representation of complex functions; aspects of boundary value problems for ordinary and partial differential equations (separation of variables, Green's functions, integral equations, variational and perturbation techniques). DFA

Algebraic Geometry, S(18), P*, L. History of Algebraic Geometry. Jean Dieudonné. Transl: Judith D. Sally. Math. Ser. Wadsworth, 1985, xii + 186 pp, $32.95. [ISBN: 0-534-03723-2] A concise survey of the evolution of algebraic geometry, "the area of mathematics where the deviation is greatest" between its intuitive roots and the abstract forms of modern research. Four of seven epochs (Greeks through Riemann) are covered quickly (in 21 pages); three others, covering works of the last hundred years, occupy the bulk of this slim volume. A final chapter surveys open problems and current work. LAS

Differential Geometry, T(17-18: 1), S, P. The General Stokes' Theorem. Helmut Grunsky. Surveys & Reference Works in Math., No. 9. Pitman, 1983, x + 113 pp, $32.95. [ISBN: 0-273-08510-7] A pro-

gressive development of Stokes' theorem in increasing generality: first in the plane, then on "pieces of space," finally on manifolds. Unusual feature: the differential of a form is initially defined analytically; calculus of alternating multilinear forms is considerably delayed. With exercises and solutions. PZ

Differential Geometry, S(18), P. Proceedings Seminar 1981-1982: Mathematical Structures in Field Theories. Ed: E.M. de Jager, H.G.J. Pijls. CWI Syllabus, V. 2. Math Centrum, 1984, iii + 217 pp, Dfl. 31 (P). [ISBN: 90-6196-278-1] A compilation of various lectures given in 1981-82 in a seminar at the University of Amsterdam in the area of gauge field theory. Some introductory material is included for the novice. Topics include quantum theory, field theory, and Yang-Mills equations. References, no index. JS

Differential Geometry, P. Differential Geometry. Ed: W. Waliszewski, G. Andrzejczak, P.G. Walczak. Banach Center Pub, V. 12. PWN, 1984, 288 pp. [ISBN: 83-01-04967-7] Partial proceedings of the 14th semester (September-December, 1979) at the Banach Center. 20 papers (2 in Russian) on two themes: global problems of Riemannian geometry, and applications of differential geometry to mathematical physics. Several are expository. PZ

Topology, T(17), S, P. Fundamentals of General Topology: Problems and Exercises. A.V. Arkhangel'skii, V.I. Ponomarev. Math. & Its Applic. D Reidel, 1984, xvi + 415 pp, $69. [ISBN: 90-277-1355-3] Problem book on set-theoretic topology. Each section begins with an explanation of basic concepts and definitions followed by an extensive list of problems which develop the theory and solutions to the problems. Problems range from routine to open research questions. Translated from Russian. BH

Control Theory, P. The Ulam Problem of Optimal Motion of Line Segments. V.A. Dubovitskij. Ser. in Math. & Eng. Optimization Software, 1985, xiii + 113 pp, $24. [ISBN: 0-911575-04-9] The author gives a closed form solution to the following problem: Among all continuous motions of an oriented line segment S in E^n from one position to another which preserves its length and for which the end-points of S lie on prescribed surfaces, find one for which the sum of the lengths of the paths swept by its endpoints is minimal. LCL

Systems Theory, S(17-18), P. Introduction to System Science. Gary M. Sandquist. Prentice-Hall, 1985, xv + 567 pp, $37.95. [ISBN: 0-13-498692-X] Application of mathematical modelling (especially linear and differential equation models) and quantitative methods to the study and analysis of systems. Based on the principle of causality with system equations (system "kernel") and functional relationships between inputs and outputs playing a central role. Applications to the sciences, as well as to economics, social sciences, humanities. RM

Probability, T(15-17: 1), S. An Introduction to Probability and Its Applications. Richard J. Larsen, Morris L. Marx. Prentice-Hall, 1985, xii + 404 pp, $28.95. [ISBN: 0-13-493453-9] Adaptable to a diverse audience: traditional topics usually found in one-semester, non-measure-theoretic undergraduate courses (random variables, expected values, special distributions), plus a final chapter on limit theorems which requires a knowledge of real analysis. High priority given to motivation; extensive examples; case studies based on real data. A companion to the authors' Introduction to Mathematical Statistics and Its Applications (TR, June-July 1981). LCL

Probability, P. Controlled Markov Processes: Time Discretization. N.M. van Dijk. CWI Tract, V. 11. Math Centrum, 1984, i + 166 pp, Dfl. 23.80 (P). [ISBN: 90-6196-280-3] First part of author's thesis. Approximates continuous-time Markov processes by discrete-time Markov processes and provides examples. MT

Probability, T(15-16: 1), S*, P, L.** Random Walks and Electric Networks. Peter G. Doyle, J. Laurie Snell. Carus Math. Mono., No. 22. MAA, 1984, xiii + 159 pp, $24. [ISBN: 0-88385-024-9] A delightful excursion among various interpretations of random walks--probability theory, linear algebra, games, electrical networks--all leading to several proofs and interpretations of Pólya's 1921 theorem that only in dimensions one and two will a random walk necessarily return to its starting point. In network terms, this is because the resistance to infinity of a lattice of resistors in dimension three (and higher) is finite. LAS

Probability, P. Lectures on Stochastic Control and Nonlinear Filtering. M.H.A. Davis. Springer-Verlag, 1984, iii + 109 pp, $7.10 (P). [ISBN: 0-387-13343-7] Two series of lecture notes, one on controlled stochastic jump processes, the other on nonlinear filtering. Parts are united by using methods of stochastic calculus to treat Markov processes. MT

Probability, T(18), S, P. Interacting Particle Systems. Thomas M. Liggett. Grund. der math. Wissenschaften, V. 276. Springer-Verlag, 1985, xv + 488 pp, $54. [ISBN: 0-387-96069-4] Intended as reference or basis for advanced graduate course. Large bibliography. Open problems after each chapter. Develops tools such as coupling and duality and applies these to study general spin systems, the stochastic Ising model, the voter model, the contact process, nearest particle systems, the exclusion process, and processes with unbounded values. BH

Statistics, T*(18: 2), S. An Introduction to Multivariate Statistical Analysis, Second Edition. T.W. Anderson. Wiley, 1984, xvii + 675 pp, $44.95. [ISBN: 0-471-88987-3] Major updating of the original edition (published in 1958), this new edition retains mathematical rigor. Augments method of maximum likelihood by alternatives such as Stein and Bayes estimators; also augments likelihood ratio tests with other invariant procedures. MT

Statistics, T(13-14: 1), S. Statistics, An Introduction. Robert N. Goldman, Joel S. Weinberg. Prentice-Hall, 1985, xii + 780 pp. [ISBN: 0-13-845918-5-01] Presupposes only high school algebra. The usual topics plus some exploratory data analysis and some discussion of MINITAB. Good examples and exercises using real data. The authors' strange notion of what should be called a definition will bother some users. FLW

Statistics, P. Lecture Notes in Statistics-26: Robust and Nonlinear Time Series Analysis. Ed: J. Franke, W. Härdle, D. Martin. Springer-Verlag, 1984, ix + 286 pp, $15 (P). [ISBN: 0-387-96102-X] Contains papers presented at a workshop at the University of Heidelberg in September 1983. Robust methods are designed to do well when distributional assumptions are false. MT

Statistics, T(17: 1), L. The Analysis of Time Series: An Introduction, Third Edition. C. Chatfield. Chapman & Hall, 1984, xiv + 286 pp, $17.95 (P). [ISBN: 0-412-26030-1] Comprehensive introduction to time series analysis, this edition is structured the same as the first two editions (First Edition, TR, June-July, 1976; Second Edition, TR, January 1982); major new feature of this edition is an appendix giving several detailed worked examples. MT

Statistics, S(15). Statistical Graphics: Design Principles and Practices. Calvin F. Schmid. Wiley, 1983, x + 212 pp, $26.50 (P). [ISBN: 0-471-87525-2] Provides guidance in selecting appropriate charts and graphs, and in drafting them informatively and effectively. Most points emphasized by examples illustrating bad chart design. MT

Statistics, T(13-14: 1, 2), S. Business Statistics, Methods and Applications. Philip G. Enns. Richard D Irwin, 1985, xxiv + 777 pp, $28.95. [ISBN: 0-256-02446-4] Some exploratory data analysis along with the usual topics and some nonparametric tests, times series, and index numbers. No calculus presupposed. FLW

Statistics, P. Lectures in Computational Statistics. Ed: J.M. Chambers, et al. Compstat Lect, V. 3. Physica-Verlag, 1984, 94 pp, $15.50 (P). [ISBN: 3-7051-0006-8] Contains two papers, "Correspondence Analysis and Gaussian Ordination" by P. Ihm and H. van Groenewoud, and "Estimating the Degrees of an Arma Model" by A. Berlinet. First paper deals with the problem of ordering a set of units on which multidimensional observations have been made. Second paper reviews available methods of estimating the unknown degrees of an auto-regressive moving average (Arma) process directly from the empirical auto-correlation function. RSK

Statistics, S(18). Testing Statistical Hypotheses: Worked Solutions. W.C.M. Kallenberg, et al. CWI Syllabus, V. 3. Math Centrum, 1984, iii + 310 pp, Dfl. 45.10 (P). [ISBN: 90-6196-280-3] Solutions to all the problems in E.L. Lehmann's well-known 1959 book Testing Statistical Hypotheses. RSK

Statistics, P. Nonparametric Density Estimation: The L_1 View. Luc Devroye, László Györfi. Wiley, 1985, xi + 356 pp, $37.95. [ISBN: 0-471-81649-9] In the Wiley Series in Probability and Mathematical Statistics. Concerned with the theory of convergence of density estimates. Includes applications to simulation, discrimination and detection. Good sets of references. RSK

Statistics, T*(15-17: 1-3), L. Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs, Second Edition. John Neter, William Wasserman, Michael H. Kutner. Richard D Irwin, 1985, xx + 1127 pp, $35.95. [ISBN: 0-256-02447-2] Extensive revision of Neter and Wasserman's 1974 First Edition (TR, December 1974). First 15 chapters have been published separately under the title Applied Linear Regression Models (TR, June 1983). Last 16 chapters contain two new chapters on nested designs, an expanded treatment of the unbalanced case in the analysis of variance, and a variety of other improvements. RSK

Statistics, T(13: 1, 2). Introductory Statistics. John W. McGhee. West, 1985, xiii + 619 pp. [ISBN: 0-314-85277-8]; Instructor's Manual with Test Bank to Accompany Introductory Statistics, v + 304 pp, (P) [ISBN: 0-314-87245-0]; Student Solutions Manual to Accompany Introductory Statistics, ix + 445 pp, (P). [ISBN: 0-314-87246-9] Fairly standard treatment. Unique features include research and case studies to illustrate applications, sample Minitab computer printouts, and library projects which involve recent journal articles using statistical techniques. RSK

Computer Programming, T(13: 1). Introduction to BASIC Programming, Second Edition. Steven L. Mandell. West, 1985, viii + 175 pp, $11.95 (P) [ISBN: 0-314-85263-8]; Instructor's Manual with Test Bank, iii + 200 pp, (P). [ISBN: 0-314-89282-6] Presents ANSI standard BASIC, but also points out variations found in microcomputer implementations. Flow charts are given for every example, although not dealt with at any length. Does not introduce any control structures (e.g., GOTO, IF-THEN, FOR-NEXT) until the fourth and fifth sections (out of ten), and does not introduce arrays until section 9. MT

Computer Programming, T(16: 1), S, P*, L. Logic Programming. Ed: K.L. Clark, S.-A. Tärnlund. APIC Stud. in Data Proc., V. 16. Academic Pr, 1984, xvii + 366 pp, $37.50. [ISBN: 0-12-175520-7] 23 important papers either from the First (1982) International Workshop on Logic Programming or outgrowths of that conference. Included are papers on the foundations of logic programming, implementation issues especially regarding PROLOG and applications to such areas as natural language processing, data bases, expert systems, and law. RWN

Computer Programming, T(13). Structured BASIC for Mini- and Microcomputers. Frank C. Lin. Reston, 1985, xix + 474 pp, (P). [ISBN: 0-8359-7127-09] A fairly traditional introduction to the BASIC programming language. The author has also included a description of non-standard features found on

many of the current micro- and mini-computers. These language extensions include structured control statements, graphics, random-access files, and additional data types. There is no discussion of non-language issues such as algorithms or problem solving. This is exclusively a language-reference book. MS

Computer Programming, T(15-16), S. FORTH. W.P. Salman, O. Tisserand, B. Toulout. Springer-Verlag, 1984, ix + 159 pp, $14 (P). [ISBN: 0-387-91256-8] FORTH is a programming language which runs on a number of well known micro- and mini-computers. It is a language very different from other more well known and popular languages, like COBOL, BASIC, or Pascal. It is a very extensible language which allows the programmer to create new statements and new data objects at will. This text gives historical background and then reviews all of the syntax and semantics of the language. At the end of the book there are a number of case studies to illustrate its power. MS

Computer Programming, T(13: 1). Problem Solving and Structured Programming in Pascal, Second Edition. Elliot B. Koffman. Addison-Wesley, 1985, xvii + 596 pp, (P). [ISBN: 0-201-11736-3] Revision of the First Edition (TR, June-July 1981) to bring it in line with ACM CS I and the AP course guidelines. Stresses problem solving through stepwise refinement, early introduction to procedures. Several case studies; sample programs, well commented, some with assertions and loop invariants. RM

Computer Programming, T?(13-14: 1). Advanced BASIC for Business. Dale E. Nelson, et al. Reston, 1985, vii + 374 pp, (P). [ISBN: 0-8359-0010-X] Following two-chapter review of elementary BASIC, covers more advanced features of BASIC-PLUS (for the PDP-11) used in business data processing: files, strings, virtual arrays. Applications to transaction and master files, sorting, billing. Long and complex examples, sample programs, and runs are difficult to read. Too few exercises. RM

Computer Programming, S, P. Forth Tools and Applications. Gary Feierbach, Paul Thomas. Reston, 1985, vi + 154 pp, (P). [ISBN: 0-8359-2091-7] This text shows how to use the Forth programming language to construct some useful support programs (i.e., tools) to help the professional programmer. The tools constructed include a memory dumper, decompiler, concordance, floating-point software, sorting, and searching routines. This is not a FORTH language reference text, and it assumes a prior knowledge of the language. MS

Software Systems, P, L. Extending the S System. Richard A. Becker, John M. Chambers. Statistics/Probability Ser. Wadsworth, 1985, viii + 166 pp, $14.95 (P). [ISBN: 0-534-05016-6] A sequel to the authors' 1984 S: An Interactive Environment for Data Analysis and Graphics (TR, February 1985) giving details on writing new functions for S, writing algorithms (in RATFOR) supporting S, and creating graphical algorithms. LAS

Computer Science, S, P. System Design with Microprocessors, Second Edition. D. Zissos. Academic Pr, 1984, viii + 191 pp, $19.50 (P). [ISBN: 0-12-781740-9] This text is a hands-on, how-to guide for building overall computer systems from existing off-the-shelf microprocessor chips. The text starts with a review of logic design and microprocessor operations. It then goes on to show how each of the components needed in an overall system can be synthesized, including interrupt control, DMA controllers, and bus controllers. The text shows explicit examples using the Intel 8080, Motorola 6800, and Motorola 6502 chips. MS

Computer Science, S(15-17), L*. Programming Languages: A Grand Tour, Second Edition. Ed: Ellis Horowitz. Computer Science Pr, 1985, ix + 758 pp, $39.95. [ISBN: 0-88175-073-5] An anthology surveying the history of programming language design. The major change in this edition is the addition of the complete "Reference Manual for the Ada Programming Language" (ANSI/MIL-STD 1815A-1983). (First Edition, TR, May 1983.) AO

Computer Science, T(15-16), P. Introduction to Natural Language Processing. Mary Dee Harris. Reston, 1985, xv + 368 pp, (P). [ISBN: 0-8359-3254-0] Natural language is any language used for communication by human beings. Getting a computer to "understand" natural language (i.e., accept it for input, produce it on output) would be an enormous advantage because it would greatly simplify the man-machine interface. This text surveys the current state of natural language understanding research and discusses the problems with getting a computer to process these languages. MS

Computer Science, S(13-15), L*. Recursion via Pascal. J.S. Rohl. Computer Sci. Texts, V. 19. Cambridge U Pr, 1984, x + 192 pp, $34.50; $14.95 (P). [ISBN: 0-521-26329-8; 0-521-26934-2] A study of recursion in procedural (rather than functional) programming. Carefully organized. Contains over 100 examples coded in Pascal. AO

Computer Science, T(17-18: 1), P, L. Fundamentals of the Average Case Analysis of Particular Algorithms. Rainer Kemp. Ser. in Computer Sci. Wiley, 1984, viii + 233 pp, $34.95. [ISBN: 0-471-90322-1] A textbook that focuses on the average case analysis of algorithms. Uses results from complex variable theory, number theory, probability theory, discrete mathematics, and combinatorics. AO

Computer Science, T(15-16: 1). Fundamentals of Operating Systems, Third Edition. A.M. Lister. Springer-Verlag, 1984, xiii + 161 pp, $13.95 (P). [ISBN: 0-387-91251-7] An updated and slightly revised edition of a standard introductory textbook. New material on monitors has been added and the chapter on file systems has been reorganized. AO

Computer Science, T(16-17: 1, 2), L. The Analysis of Algorithms. Paul Walton Purdom, Jr., Cynthia A. Brown. Holt, Rinehart & Winston, 1984, xv + 540 pp. [ISBN: 0-03-072044-3] A textbook for an

elementary/intermediate course in algorithm analysis. The required mathematics (beyond calculus) is largely self-contained. AO

Computer Science, P. Applied Probability--Computer Science: The Interface, Volume I. Ed: Ralph L. Disney, Teunis J. Ott. Progress in Comp. Sci., No. 2. Birkhauser Boston, 1982, xxvii + 504 pp, $39.95. [ISBN: 3-7646-3067-8] The first of two volumes containing the proceedings of the ORSA/TIMS special interest conference held in January 1981, at Florida Atlantic University, Boca Raton. AO

Applications. Transactions of the Second Army Conference on Applied Mathematics and Computing. US Army Research Office (PO Box 12211, Research Triangle Park, NC 27709), 1985, xxii + 952 pp, (P). Proceedings of the conference held at Resselaer Polytechnic Institute in May 1984. Papers from a diverse range of research areas including robotics, continuum mechanics, computational methods, and constitutive equations for high strain rate problems. AO

Applications. Teaching and Applying Mathematical Modelling. Ed: J.S. Berry, et al. Ser. in Math. & Its Applic. Halsted Pr, 1984, xvii + 491 pp, $75. [ISBN: 0-470-20079-0] Papers from the "First International Conference on the Teaching of Mathematical Modeling" held at the University of Exeter. AO

Applications, P, L. Foundations of Computer Music. Ed: Curtis Roads, John Strawn. MIT Pr, 1985, xv + 712 pp, $50. [ISBN: 0-262-18114-2] Reprints of 36 articles from Volumes 1-3 (1977-79) of Computer Music Journal, covering seminal papers on digital sound synthesis, hardware, software systems, and perception. An accompanying volume Computer Music Tutorial provides introductory definitions for students. LAS

Applications, S(13), L*. Mathematics in Sport. M. Stewart Townend. Ser. in Math. & Its Applic. Halsted Pr, 1984, 202 pp, $19.95 (P). [ISBN: 0-470-20082-4] An elementary analysis of the physics of sports--running, jumping, throwing--based on considerable empirical data, simple algebra, and differential equations. A good source of enrichment examples for beginning college mathematics, although the British author fails to deal with American football! LAS

Applications (Actuarial Science), P. Risk Theory: The Stochastic Basis of Insurance, Third Edition. R.E. Beard, T. Pentikäinen, E. Pesonen. Chapman & Hall, 1984, xvii + 408 pp, $24 (P); $49.95. [ISBN: 0-412-25980-X; 0-412-24260-5] Introduction to risk theory devoted primarily to practical applications, as opposed to theoretical approach of other books on risk theory. Current edition adds new developments to the material from the Second Edition (First Edition, TR, April 1970; Second Edition, TR, February 1978). MT

Applications (Economics), S(17-18), P*, L.** Handbook of Econometrics. Ed: Zvi Griliches, Michael D. Intriligator. Volume I, Handbooks in Econ., B. 2. Elsevier Sci, 1983, xxvii + 771 pp, $66 [ISBN: 0-444-86185-8]; Volume II, 1984, xxvi + 686 pp, $65. [ISBN: 0-444-86186-6] The first two of three volumes of a planned 35-chapter survey of econometrics intended to serve as a reference and teaching supplement. Volume I contains twelve chapters covering mathematical and statistical methods, econometric models, estimation, and computation. Volume II contains twelve additional chapters on testing, time series, and special topics. This is the second multi-volume Handbook, the first being the Handbook of Mathematical Economics (TR, Volume I, March 1982; Volume II, October 1982). LAS

Applications (Physics), P. Applications of Group Theory in Physics and Mathematical Physics. Ed: Moshe Flato, Paul Sally, Gregg Zuckerman. Lect. in Appl. Math., V. 21. AMS, 1985, xii + 420 pp, $70. [ISBN: 0-8218-1121-5] Papers from the Summer Seminar held in July 1982 in Chicago. Major subject areas represented are supersymmetry and supergravity, representations of noncompact Lie groups, and Kac-Moody algebras and nonlinear theories. AO

Applications (Physics), P. Phase Transformations and Material Instabilities in Solids. Ed: Morton E. Gurtin. Academic Pr, 1984, ix + 217 pp, $17. [ISBN: 0-12-309770-3] The invited talks from a conference held at the University of Wisconsin, Madison, in October 1983. Topics include general theories of phase transitions, equilibrium shapes of surfaces, morphological instabilities and dendrite formation, shock-induced phase transitions, and related results in the calculus of variations. AO

Reviewers

DFA: David F. Appleyard, Carleton; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; RD: Roger Day, St. Olaf; JD-B: John Dyer-Bennet, Carleton; SG: Steven Galovich, Carleton; BH: Bruce Hanson, St. Olaf; PH: Paul Humke, St. Olaf; KK: Kenneth Kaminsky, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; AM: Alan Magnuson, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; LN: Linda Ness, Carleton; AO: Arnold Ostebee, St. Olaf; MS: Michael Schneider, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MT: Michael Tveite, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

The standard proof of Jensen's formula ultimately involves showing that

$$\int_0^{2\pi} \log|1 - e^{i\theta}|\,d\theta = 0,$$

which is then established by means of the residue theorem (see, e.g., [1], [2], [3], and [4]). But residues are unnecessary—the integral is in fact completely elementary, as the following calculations show.

Since $|1 - e^{i\theta}| = 2\sin\frac{\theta}{2}$, we need only verify that

$$\int_0^{\pi} \log\sin x\,dx = -\pi\log 2.$$

Call the left side I and write $\sin x = 2\sin\frac{x}{2}\cos\frac{x}{2}$. Then

$$I = \pi\log 2 + \int_0^{\pi}\log\sin\frac{x}{2}\,dx + \int_0^{\pi}\log\cos\frac{x}{2}\,dx.$$

Replacing $\frac{x}{2}$ by t in the first integral and by $\frac{\pi}{2} - t$ in the second, we find

$$\begin{aligned} I &= \pi\log 2 + 4\int_0^{\pi/2}\log\sin t\,dt \\ &= \pi\log 2 + 2I, \end{aligned}$$

and the result follows.

References

1. L. V. Ahlfors, Complex Analysis, 3rd ed., McGraw-Hill, 1979.
2. J. B. Conway, Functions of One Complex Variable, 2nd ed., Springer-Verlag, 1978.
3. N. Levinson and R. M. Redheffer, Complex Variables, Holden-Day, 1970.
4. W. Rudin, Real and Complex Analysis, 2nd ed., McGraw-Hill, 1973.

ANTISOCIAL SUBCOVERS OF SELF-CENTERED COVERINGS

STEVEN G. KRANTZ* AND T. D. PARSONS
Department of Mathematics, Pennsylvania State University, University Park, PA 16802

1. Introduction. It is a remarkable and relatively recent discovery that fundamental results in measure theory are logically equivalent to certain covering properties of balls in Euclidean space. For instance, consider the Lebesgue Differentiation Theorem, which says that if f is a Lebesgue integrable function on $\mathbb{R}^n$, m is Lebesgue measure on $\mathbb{R}^n$, and $B(x,r)$ is the ball $\{t \in \mathbb{R}^n: |x - t| < r\}$, then

$$\lim_{r\to 0^+}\frac{1}{m(B(x,r))}\int_{B(x,r)} f(t)\,dt = f(x)$$

for almost every $x \in \mathbb{R}^n$. This result is logically equivalent to the following covering theorem:

THEOREM A. *Let K be a compact subset of $\mathbb{R}^n$ and let $\mathscr{C} = \{B(x_i, r_i): i \in I\}$ be a collection of balls which cover K. Then there is a finite subcollection $\{B(x_j, r_j): j \in J\} \subseteq C$, whose members are pairwise disjoint, such that $K \subset \bigcup_{j\in J} B(x_j, 3r_j)$.* (See [3], [1] for details on these results.)

Thus covering theorems like Theorem A play a central role in modern real and harmonic analysis. In this note we prove a covering theorem of a slightly different nature and then derive a surprising theorem about partitions of covers. Theorem A holds in any metric space. See [2] and references therein.

*Research supported in part by a grant from the National Science Foundation.

2. Antisocial coverings. Let X be a metric space. For $r > 0$ and $x \in X$, let $B(x, r) = \{t \in X: \operatorname{dist}(x, t) < r\}$ be the open ball of radius r centered at x. A *self-centered covering* $\mathscr{C}$ of a non-empty subset S of X is a collection $\mathscr{C} = \{B(x, r_x): x \in S\}$ of open balls, with one ball of $\mathscr{C}$ centered at each point of S. An *antisocial family* $\mathscr{A}$ is a family of balls such that if $B(x, r)$ and $B(y, s)$ are two different balls in $\mathscr{A}$, then $y \notin B(x, r)$ and $x \notin B(y, s)$; that is, no ball in F contains the center of another. Our theorem is:

THEOREM B. *Let X be a metric space and let K be a non-empty compact subset of X. Let $\mathscr{C} = \{B(x, r_x): x \in K\}$ be a self-centered covering of K. Suppose that the radius mapping $x \to r_x$ is continuous on K. Then there is an antisocial family $\mathscr{A} \subseteq \mathscr{C}$ such that $\mathscr{A}$ covers K.*

Proof. By the compactness and continuity assumptions, the radius function $r(x) = r_x$ attains its maximum on each non-empty closed subset of K. Let $K_0 = K$ and let x_0 be any location of this maximum on K_0. Recursively define x_{n+1} to be any point of $K_{n+1} \equiv K \setminus \bigcup_{i=0}^{n} B(x_i, r(x_i))$, where $r(x)$ attains its maximum on K_{n+1}. (If $K \setminus \bigcup_{i=0}^{n} B(x_i, r(x_i)) = \varnothing$, then terminate the construction and do not define x_{n+1}.)

If $i < j$, then $x_j \notin B(x_i, r(x_i))$, so that $\operatorname{dist}(x_i, x_j) \geqslant r(x_i) \geqslant r(x_j)$, hence also $x_i \notin B(x_j, r(x_j))$. We claim that the sequence must terminate after finitely many terms. Suppose not. Then the sequence $\{x_n\}$ has an accumulation point $x \in K$, so some subsequence $\{x_{n_p}\}$ converges to x, and the non-increasing sequence $\{r(x_{n_p})\}$ converges to $r(x) > 0$. It is then easy to see that, for large p, each of the balls $B(x_{n_p}, r(x_{n_p}))$ contains the ball $B(x, r(x)/2)$, contradicting the antisociality of the collection $\{B(x_n, r(x_n))\}$.

Therefore for some integer n, $K \subseteq \bigcup_{i=0}^{n} B(x_i, r(x_i))$, and $\mathscr{A} = \{B(x_i, r(x_i)): 0 \leqslant i \leqslant n\}$ gives the desired antisocial subcover of $\mathscr{C}$. $\square$

REMARK. We note that any covering of a compact set K by an antisocial family whose centers belong to K must be finite—because each center is in just one ball of the family.

3. Very antisocial partitions. In $\mathbb{R}^n$, each finite antisocial family $\mathscr{A}$ of open balls can be partitioned into subfamilies $\mathscr{A}_1, \mathscr{A}_2, \ldots, \mathscr{A}_m$ where $m \leqslant f(n)$ for some constant $f(n)$ depending only on the dimension, such that each subfamily $\mathscr{A}_i$ is very antisocial indeed—the balls in $\mathscr{A}_i$ are pairwise disjoint! This result is stated formally as Theorem C below. We now develop a simple proof of this fact.

LEMMA 1. *Let K be a compact subset of a metric space X, and let $\lambda > 0$. Then there exists a least positive integer $N = N(K, \lambda)$ such that for every subset $\{x_1, \ldots, x_n\}$ of K with $\operatorname{dist}(x_i, x_j) \geqslant \lambda$ for $i \neq j$, we have $n \leqslant N$.*

Proof. The covering $\{B(x, \frac{1}{2}\lambda): x \in K\}$ of K has a finite subcover $\mathscr{U} = \{B(y_i, \frac{1}{2}\lambda): i = 1, \ldots, s\}$. Let $N(K, \lambda) = s$. Suppose that $x_1, \ldots, x_n \in K$ and $\operatorname{dist}(x_i, x_j) \geqslant \lambda$ for $i \neq j$. If $n > N(K, \lambda)$, then two of the points, x_p and x_q, must lie in the same ball $B(y_i, \frac{1}{2}\lambda)$ of $\mathscr{U}$. But then $\operatorname{dist}(x_p, x_q) < \lambda$, a contradiction. Therefore $n \leqslant N(K, \lambda)$. $\square$

LEMMA 2. *Let Σ denote the sector $\{re^{i\theta}: 0 \leqslant r \leqslant \infty, 0 \leqslant \theta \leqslant \pi/6\}$ in $\mathbb{R}^2$, and put $D = \{w \in \Sigma: |w| \geqslant 3\}$. If $x, y \in D$ and each of $B(x, r), B(y, s)$ intersects $B(0,1)$, then $|x - y| < \max(r, s)$.*

Proof. We first need some simple inequalities:

(i) $(a-1)^2 - (2-\sqrt{3})a^2 \geqslant 0$ if $a \geqslant 3$;

(ii) $(b-1)^2 - (a^2 - \sqrt{3}\,ab + b^2) \geqslant 0$ if $b \geqslant a \geqslant 3$.

The derivative of the expression in (i) is positive when $a \geqslant 3$, and since (i) holds when $a = 3$, (i) is true for every $a \geqslant 3$. Similarly, the derivative of (ii) with respect to b is positive when $b \geqslant a \geqslant 3$, and (ii) holds when $b = a \geqslant 3$ by (i).

Now write $x = ae^{i\theta}$, $y = be^{i\varphi}$; we may assume $a \leqslant b$. The assumption $B(y, s) \cap B(0,1) \neq \varnothing$ clearly implies $s > b - 1$, so it suffices to show $|x - y|^2 \leqslant (b-1)^2$. By the law of cosines,

$$|x - y|^2 = a^2 - 2ab\cos(\varphi - \theta) + b^2,$$

which is less than or equal to $a^2 - \sqrt{3}\,ab + b^2$ because

$$\cos(\varphi - \theta) \geqslant \cos \pi/6 = \sqrt{3}/2.$$

The inequality $|x - y|^2 \leqslant (b-1)^2$ thus follows from (ii). □

THEOREM C. *Let n be a positive integer. Then there exists a least positive integer $f(n)$ such that every finite antisocial family $\mathscr{A}$ of open balls in $\mathbb{R}^n$ can be partitioned into no more than $f(n)$ subfamilies in each of which the balls are pairwise disjoint.*

Proof. The case $n = 1$ is trivial, and $f(1) = 2$. Let $n > 1$, and set $\delta = |e^{i\pi/12} - 1|$. Let S be the unit sphere in $\mathbb{R}^n$. Let $\mathscr{C} = \{B(x,\delta): x \in S\}$. Applying Theorem B to the covering $\mathscr{C}$ of the compact set S, with the constant radius mapping $r(x) = \delta$, we obtain an antisocial subfamily $\mathscr{C}' = \{B(y_i,\delta): i = 1,2,\dots,m\}$ which covers S. If $1 \leqslant i < j \leqslant m$, then $y_j \notin B(y_i,\delta)$ so $\operatorname{dist}(y_i, y_j) \geqslant \delta$. Thus $m \leqslant N(S,\delta)$.

Let O be the origin in $\mathbb{R}^n$, and for $i \in \{1,2,\dots,m\}$ let V_i be the cone determined by O and $B(y_i,\delta) \cap S$. That is,

$$V_i = \{ry: y \in B(y_i,\delta) \cap S, 0 \leqslant r < \infty\}.$$

Clearly, $V_1 \cup \cdots \cup V_m = \mathbb{R}^n$ since $\mathscr{C}'$ covers S. For $i \in \{1,2,\dots,m\}$, let

$$W_i = \{x \in V_i: \operatorname{dist}(0,x) \geqslant 3\},$$

and let

$$W = \{x \in \mathbb{R}^n: \operatorname{dist}(0,x) \leqslant 3\}.$$

Note that $W \cup W_1 \cup \cdots \cup W_m = \mathbb{R}^n$. Let $p = N(W,1)$. We shall show that $f(n) \leqslant m + p$; that is, $f(n) \leqslant N(S,\delta) + N(W,1)$.

Let $\mathscr{A}$ be any finite antisocial family of open balls in $\mathbb{R}^n$. Choose a ball B_1^1 of largest radius in $\mathscr{A}$. If possible, choose B_2^1 to be a ball of largest radius disjoint from B_1^1 in $\mathscr{A}$, then B_3^1 a ball of largest radius disjoint from both B_1^1 and B_2^1, and so forth. We obtain a family $\mathscr{A}_1$ of pairwise disjoint elements of $\mathscr{A}$, such that either $\mathscr{A} \setminus \mathscr{A}_1 = \varnothing$ or else $\mathscr{A} \setminus \mathscr{A}_1 \neq \varnothing$ and every element of $\mathscr{A} \setminus \mathscr{A}_1$ intersects some element of $\mathscr{A}_1$. In the latter case, we start a new subfamily $\mathscr{A}_2$ by choosing B_1^2 of largest radius from $\mathscr{A} \setminus \mathscr{A}_1$, $B_2^2 \in \mathscr{A} \setminus (\mathscr{A}_1 \cup \{B_1^2\})$ of largest possible radius disjoint from B_1^2, etc. Continuing in this manner, we define a sequence $\mathscr{A}_1, \mathscr{A}_2, \dots$ of pairwise disjoint non-empty subfamilies of $\mathscr{A}$ such that each $\mathscr{A}_i$ consists of pairwise disjoint balls and such that each element B_k^i of $\mathscr{A}_i$ $(i > 1)$ intersects some element of $\mathscr{A}_j$ for every $j < i$. Since $\mathscr{A}$ is finite, for some least positive integer t we have $\mathscr{A} = \mathscr{A}_1 \cup \mathscr{A}_2 \cup \cdots \cup \mathscr{A}_t$.

We claim that $t \leqslant m + p$. Suppose not. Let $q = m + p$ and choose $B_0 = B(x_0, r_0) \in \mathscr{A}_{q+1}$. For each $i \in \{1,2,\dots,q\}$ let $B_i = B(x_i, r_i) \in \mathscr{A}_i$ be the first ball which intersects B_0 in the sequence $B_1^i, B_2^i, \dots$ of elements generating $\mathscr{A}_i$. Note that $r_i \geqslant r_0$, since otherwise B_0 would have been chosen instead of B_i in the construction of $\mathscr{A}_i$.

Without loss of generality, we may assume that $x_0 = 0$ and $r_0 = 1$ (otherwise, just replace $\mathscr{A}$ by the "normalized" family $\{B(y - x_0, \rho/r_0): B(y,\rho) \in \mathscr{A}\}$). Since each $r_i \geqslant r_0 = 1$ and since $i \neq j$ implies that $\operatorname{dist}(x_i, x_j) \geqslant \max(r_i, r_j)$ because $\mathscr{A}$ is antisocial, we have that at most $p = N(W,1)$ of the points $x_0, x_1, \dots, x_q$ are in W (and one of these must be x_0, since $x_0 = 0 \in W$). Suppose that two points $x_j, x_k \in W_i$ for some $i \in \{1,2,\dots,m\}$. Then we may choose a plane Π containing x_0, x_j, x_k and this plane intersects the closure of the cone V_i in a copy of the sector Σ appearing in Lemma 2 because of our choice of δ. By Lemma 2, since x_j and x_k must lie inside the region D of the lemma, we get $\operatorname{dist}(x_j, x_k) < \max(r_j, r_k)$, which is impossible because $\mathscr{A}$ is antisocial. Therefore each W_i contains at most one of the points $x_1, x_2, \dots, x_q$. Since $\mathbb{R}^n = W \cup W_1 \cup \cdots \cup W_m$, we conclude that

$$q + 1 = |\{x_0, x_1, \dots, x_q\}| \leqslant p + m = q,$$

a contradiction. Therefore $t \leqslant m + p$, as claimed, so $f(n) \leqslant m + p \leqslant N(S,\delta) + N(W,1)$. □

4. Remarks and problems. It appears that our hypothesis in Theorem B that the radius mapping $x \to r(x)$ be continuous can be weakened considerably, or perhaps be dropped entirely. Our proof needed only that $r(x)$ attain a maximum on each compact subset of K; therefore we could replace the continuity assumption by upper semi-continuity, or by the hypothesis that $r(x)$ take on only finitely many values. If X is also a measure space and if we require that the antisocial subfamily cover all but a subset of K of measure zero, then we need only assume $r(x)$ to be measurable. Perhaps Theorem B holds with no restrictions on the radius mapping, but we cannot yet prove or disprove this.

PROBLEM 1. Does every self-centered covering of a compact metric space have an antisocial subcover?

Determination of the exact values of the function $f(n)$ of Theorem C seems to be quite difficult. The upper bound given in our proof is not very sharp even for the case $n = 2$. By more careful analysis, still based on angular sectors, we can show that $f(2) \leqslant 20$, but surely $f(2)$ is much smaller than 20.

PROBLEM 2. Compute $f(2)$.

References

1. A. Cordoba and R. Fefferman, On differentiation of integrals, Proc. Nat. Acad. Sci. U.S.A., 74 (1977) 2211–2213.
2. H. Federer, Geometric Measure Theory, Springer-Verlag, Berlin, 1970.
3. E. M. Stein, Singular Integrals and Differentiability Properties of Functions, Princeton University Press, 1970.

LINEAR CONVERGENCE AND THE BISECTION ALGORITHM

EDWIN H. KAUFMAN, JR., AND TERRY D. LENKER
Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859

1. Introduction. The bisection algorithm is often the first algorithm presented in a numerical analysis course to find the (assumed) unique zero of $f \in C[0,1]$. It is well known that the iterates produced by this algorithm converge to the zero, and the number of iterations required to insure a desired degree of accuracy can be calculated. It is desirable to end the discussion with an analysis of the rate of convergence of the algorithm. Unfortunately, the current literature ([**1**], [**2**], [**3**], [**5**], [**7**]) is divided as to whether or not the rate of convergence of the bisection algorithm is linear, and does not even agree on the definition of linear convergence. We will analyze the convergence properties of the bisection algorithm.

Suppose $f \in C[0,1]$ has a unique zero x in $(0,1)$ and $f(0) \cdot f(1) < 0$. The bisection algorithm produces a sequence of iterates x_k in the following manner. Set $x_1 = 1/2$, $a_1 = 0$, and $b_1 = 1$, and suppose we have found x_i, a_i, and b_i for some $i \geqslant 1$. Then

(i) if $f(x_i) = 0$, the algorithm terminates;

(ii) if $f(a_i) \cdot f(x_i) < 0$, let

$$x_{i+1} = (a_i + x_i)/2 = x_i - (1/2)^{i+1}, a_{i+1} = a_i \quad \text{and} \quad b_{i+1} = x_i \text{ (see Fig. 1)};$$

(iii) if $f(a_i) \cdot f(x_i) > 0$, let

$$x_{i+1} = (x_i + b_i)/2 = x_i + (1/2)^{i+1}, a_{i+1} = x_i \quad \text{and} \quad b_{i+1} = b_i \text{ (see Fig. 2)}.$$

The algorithm either terminates with $f(x_n) = 0$ for some n, or the sequence $\{x_k\}_{k=1}^{\infty}$ converges to x with $|x_k - x| \leqslant (1/2)^k$ for all k.

We compare the rate of convergence of the bisection algorithm with three different notions of linear convergence which appear in the literature. For a comprehensive treatment of rates of

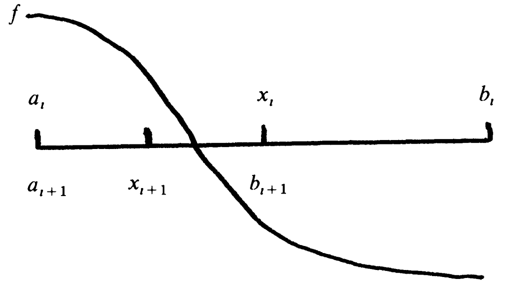


FIG. 1

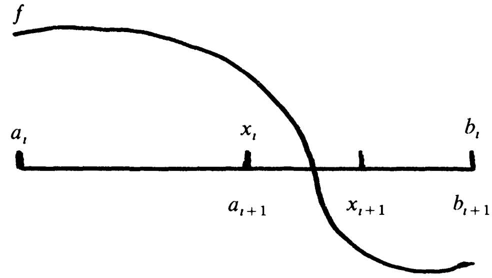


FIG. 2

convergence of algorithms we recommend [4]. For $k \geqslant 1$ let $e_k = x - x_k$ denote the error in using the kth approximation to x obtained from some algorithm.

DEFINITION. An algorithm converges *linearly* if there exists a positive constant $K < 1$ such that $|e_{k+1}| \leqslant K|e_k|$ for all sufficiently large k. K is called the *convergence factor*.

A second common definition of linear convergence replaces the condition above by the condition $\lim_{k\to\infty}(|e_{k+1}|/|e_k|) < 1$; although in general this is a stronger condition (since the limit might fail to exist), the proof of Theorem 2 below shows that in the case of the bisection algorithm, the two conditions are equivalent if the algorithm does not terminate. We also note that, in general, linear convergence of an algorithm implies what we call *geometric convergence*; that is, $\overline{\lim}_{k\to\infty}|e_k|^{1/k} < 1$.

For each $x \in (0,1)$ there are uncountably many functions f which satisfy $f(0)\cdot f(1) < 0$ and for which x is a simple zero. Henceforth we describe the rate of convergence of the bisection algorithm for certain subsets of $(0,1)$ rather than for the associated classes of functions. In particular, for all $x \in (0,1)$ for which the bisection algorithm does not terminate, the bisection algorithm possesses geometric convergence since $\overline{\lim}_{k\to\infty}|e_k|^{1/k} = 1/2$. In the next section we will characterize those $x \in (0,1)$ for which the bisection algorithm also possesses linear convergence.

2. Convergence Properties of the Bisection Algorithm. The kth iterate x_k of the bisection algorithm is given by

$$x_k = 1/2 + \sum_{i=2}^{k} s_i(1/2)^i,$$

where each $s_i = 1$ or -1. This motivates the following theorem.

THEOREM 1. *Every $x \in (0,1)$ has a unique representation of the form $x = \sum_{i=1}^{k} s_i(1/2)^i$ if $x = p/2^k$ is a dyadic rational, or $x = \sum_{i=1}^{\infty} s_i(1/2)^i$ otherwise, where $s_1 = 1$ and $s_i = 1$ or -1 for $i > 1$.*

Proof. It is known [6, p. 156] that every $x \in (0,1)$ has a unique binary representation $x = \sum_{i=1}^{\infty} c_i(1/2)^i$, where each $c_i = 0$ or 1 and $c_i = 1$ from some point on is not allowed (thus any dyadic rational $p/2^k$ can be expressed as a terminating series $\sum_{i=1}^{k} c_i(1/2)^i$).

The results above follow immediately from setting $s_{i+1} = 2c_i - 1$ for all i.

We now prove a lemma which is the key to our main results.

LEMMA. *Suppose the bisection algorithm does not terminate at x_{k+2} or sooner, where $k \geqslant 1$. Suppose also that $s_{k+1} \neq s_{k+2} = s_{k+3}$. Then $|e_{k+1}| > |e_k|$.*

Proof. Without loss of generality, we may assume

$$s_{k+1} = 1 \quad \text{and} \quad s_{k+2} = s_{k+3} = -1,$$

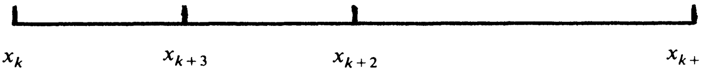


FIG. 3

so the locations of x_k, x_{k+1}, x_{k+2} and x_{k+3} are shown in Fig. 3.

Since $a_{k+3} = x_k$, $b_{k+3} = x_{k+2}$ and the algorithm did not terminate at x_{k+2} or earlier, we must have $x \in (x_k, x_{k+2})$. Thus $|x_{k+1} - x| > |x_k - x|$.

THEOREM 2. *Let $x \in (0,1)$ have the non-terminating representation $x = \sum_{i=1}^{\infty} s_i(1/2)^i$. The bisection algorithm converges linearly to x if and only if the s_i's are alternately plus and minus one from some point on.*

Proof. The hypotheses imply that there are arbitrarily large indices i with $s_i = -1$ and arbitrarily large indices i with $s_i = 1$. If the s_i's do not alternate from some point on, then there are arbitrarily large indices k with $s_{k+1} \neq s_{k+2} = s_{k+3}$, so by the lemma $|e_{k+1}| > |e_k|$. Thus the algorithm does not converge linearly.

Conversely, suppose the s_i's alternate from some point on. Then for some j,

$$s_j = s_{j+2} = s_{j+4} = \cdots \neq s_{j+1} = s_{j+3} = \cdots .$$

For $l \geqslant j$,

$$|x_{l+1} - x| = \left| \sum_{i=l+2}^{\infty} s_i(1/2)^i \right| = \frac{1}{2} \left| \sum_{i=l+1}^{\infty} s_i(1/2)^i \right| = \frac{1}{2}|x_l - x|,$$

which implies the algorithm converges linearly with convergence factor $1/2$.

We can now prove the main theorem.

THEOREM 3. *For $x \in (0,1)$, there are exactly three possibilities:*

(i) *x is a dyadic rational number of the reduced form $p/2^k$ for integers $p > 0$ and $k > 0$ if and only if the bisection algorithm terminates;*

(ii) *x is a rational number of the reduced form $p/(3 \cdot 2^k)$ for integers $p > 0$ and $k \geqslant 0$ if and only if the bisection algorithm does not terminate and converges linearly with convergence factor $1/2$;*

(iii) *x is not one of the reduced forms above if and only if the bisection algorithm does not terminate, and does not converge linearly.*

Proof. (i) Suppose the algorithm terminates at x_k. Then

$$x = x_k = \sum_{i=1}^{k} s_i(1/2)^i = \left(\sum_{i=1}^{k} 2^{k-i} s_i \right) / 2^k = p/2^k,$$

where p must be an odd integer. On the other hand suppose $x = p/2^k$ in reduced form. Then by Theorem 1 x has the unique representation $x = \sum_{i=1}^{k} s_i(1/2)^i$, so the bisection algorithm terminated at x_k.

(ii) First suppose the algorithm does not terminate and converges linearly. By Theorem 2, there exists $j > 1$ such that the s_i's alternate in sign for $i \geqslant j$. Thus we have

$$x = \sum_{i=1}^{j-1} s_i(1/2)^i + \sum_{i=j}^{\infty} s_i(1/2)^i = l/2^{j-1} + s_j(1/2)^j/(1-(-1/2)) = (3l + s_j)/(3 \cdot 2^{j-1}),$$

where l is an odd integer. So, in reduced form, $x = p/(3 \cdot 2^k)$ for some $k \leqslant j - 2$.

Now suppose $x = p/(3 \cdot 2^k)$ in reduced form, so we may write $2p = 3m + \delta$, where $\delta = \pm 1$ and m is odd. Thus,

$$x = \frac{m}{2^{k+1}} + \delta \cdot \frac{1}{2^{k+2}} \cdot \frac{2}{3} = \sum_{i=1}^{k+1} s_i(1/2)^i + \delta \cdot \frac{1}{2^{k+2}} \sum_{i=0}^{\infty} (-1)^i \left(\frac{1}{2}\right)^i$$
$$= \sum_{i=1}^{k+1} s_i \left(\frac{1}{2}\right)^i + \sum_{i=k+2}^{\infty} \delta \cdot (-1)^{i+k+2} \left(\frac{1}{2}\right)^i,$$

where $\sum_{i=1}^{k+1} s_i(\frac{1}{2})^i$ is the expansion of the dyadic rational $m/2^{k+1}$ given by Theorem 1. Thus the bisection algorithm applied to x does not terminate, and converges linearly by Theorem 2.

(iii) This follows from (i) and (ii) by exhaustion.

EXAMPLE. Let $x = 5/12 = 5/(3 \cdot 2^2)$. By Theorem 3 the bisection algorithm does not terminate and converges linearly. The binary representation of 5/12 is .01101010.... Hence

$$\{s_i\}_{i=1}^{\infty} = \{1, -1, 1, 1, -1, 1, -1, 1, -1, \dots\}$$

and we have

$$\frac{5}{12} = \frac{1}{2} - \frac{1}{4} + \frac{1}{8} + \frac{1}{16} - \frac{1}{32} + \frac{1}{64} - \frac{1}{128} + \cdots.$$

EXAMPLE. Let $x = 3/5$. The bisection algorithm does not terminate and does not converge linearly. We have

$$\frac{3}{5} = \frac{1}{2} + \frac{1}{4} - \frac{1}{8} - \frac{1}{16} + \frac{1}{32} + \frac{1}{64} - \cdots,$$

so our lemma implies that $|x_{2l} - 3/5| > |x_{2l-1} - 3/5|$ for all $l \geqslant 1$.

References

1. K. E. Atkinson, An Introduction to Numerical Analysis, Wiley, New York, 1978.
2. R. L. Burden, J. D. Faires, and A. C. Reynolds, Numerical Analysis, 2nd Ed., Prindle, Weber and Schmidt, Boston, 1981.
3. E. Isaacson and H. B. Keller, Analysis of Numerical Methods, Wiley, New York, 1966.
4. J. M. Ortega and W. C. Rheinboldt, Iterative Solutions of Non-linear Equations in Several Variables, Academic Press, New York, 1970.
5. S. M. Pizer, Numerical Computing and Mathematical Analysis, Science Research Associates, Chicago, 1975.
6. R. R. Stoll, Set Theory and Logic, Foresman, San Francisco, 1963.
7. J. S. Vandergraft, Introduction to Numerical Computations, Academic Press, New York, 1978.

A GEOMETRIC INTERPRETATION FOR THE EULERIAN NUMBERS

IBRAHIM A. SALAMA AND LAWRENCE L. KUPPER
Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27514

The Eulerian numbers $A(n, k)$ are defined by the expression

$$A(n,k) = \sum_{j=0}^{k} (-1)^j \binom{n+1}{j} (k-j)^n,$$

and they have the following well-known permutational interpretation (see [**1**], Chap. VI). Let S_n denote the set of permutations over $\{1, 2, \dots, n\}$. A permutation $\sigma = [\sigma(1), \dots, \sigma(n)] \in S_n$ is said to induce a rise (or a fall) in $i = 1, \dots, n-1$ if $\sigma(i) < \sigma(i+1)$ [or $\sigma(i) > \sigma(i+1)$]. Let R_σ be the number of rises in σ; then the number of permutations of S_n with k rises is given by $A(n, k+1)$. In this note we provide the following geometrical characterization of $A(n, k)$.

Consider the set $I^n = [0,1]^n$. Let $P_0, \dots, P_n$ be the family of $(n+1)$ planes in R^n such that:

(i) The normal to P_0 is $(1, \dots, 1)'$.

(ii) P_k is parallel to P_0, $k = 1, \dots, n$.
(iii) $P_k \cap I^n \neq \phi$, $k = 0, 1, \dots, n$.
(iv) The distance of P_k from the origin is $k/\sqrt{n}$, $k = 0, 1, \dots, n$.

Further, let S_{nk} be that portion of I^n bounded by P_{k-1} and P_k, and let $V(S_{nk})$ denote the volume of S_{nk}.

Then the following result can be proved.

THEOREM. $n!V(S_{nk}) = A(n,k)$.

Proof. Let $X_1, \dots, X_n$ be n independent and identically distributed uniform random variables defined over $[0,1]$. Let

$$G(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}_n.$$

Then the probability density function of $\overline{X}_n$ takes the form

$$f_{\overline{X}_n}(x) = \frac{n^n}{(n-1)!} \sum_{j=0}^{n} (-1)^j \binom{n}{j} \left[c\left(x - \frac{j}{n}\right) \right]^{n-1}, \quad 0 \leqslant x \leqslant 1,$$

where $c(u) = 0$ if $u < 0$, and $c(u) = u$ if $u \geqslant 0$ (See [2], pp. 257–259 for a proof of this result.)

Elementary integration then yields

$$pr\left\{ \frac{(k-1)}{n} < \overline{X}_n < \frac{k}{n} \right\} = \int_{(k-1)/n}^{k/n} f_{\overline{X}_n}(x)\, dx$$

$$= \frac{1}{n!} \sum_{j=0}^{k} (-1)^j \binom{n+1}{j} (k-j)^n = A(n,k)/n!.$$

Finally, since $S_{nk} = \{(X_1, \dots, X_n) | (k-1)/n < G(X_1, \dots, X_n) < k/n\}$, the theorem follows.

Note that the result remains true (with some minor modifications) if we replace $(1, \dots, 1)'$ by any vector parallel to a main diagonal of I^n.

References

1. L. Comtet, Advanced Combinatorics, Reidel, 1974.
2. M. G. Kendall and A. Stuart, The Advanced Theory of Statistics, 3rd ed., Hafner, 1969.

THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

Material for this department should be sent to Professor Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.

THE CONTRACTION MAPPING LEMMA AND THE INVERSE FUNCTION THEOREM IN ADVANCED CALCULUS

LANCE D. DRAGER AND ROBERT L. FOOTE
Department of Mathematics, Texas Tech University, Lubbock, TX 79409

In an undergraduate advanced calculus course, one of the important theorems a student sees is

the inverse function theorem. The most elegant and general proof of this uses the contraction mapping lemma (see [1]). Unfortunately, time does not usually permit one to go into the theory of metric spaces, Cauchy sequences, completeness, etc., necessary for the proof of the general contraction mapping lemma.

The purpose of this note is to give a simple proof of the contraction mapping lemma in $\mathbb{R}^n$ based on a principle familiar to calculus students.

MAX / MIN PRINCIPLE. *A continuous, real-valued function defined on a non-empty, closed, bounded subset of* $\mathbb{R}^n$ *attains its maximum and minimum on that set.*

A proof of this, of course, requires many of the topics one is forced to omit. We feel, however, that our use of this principle to prove the contraction mapping lemma is pedagogically valuable. Early in a calculus course one typically invokes the max/min principle (usually without proof) when discussing max/min problems, and students seem to be reasonably comfortable with it. Its application to the inverse function theorem via the contraction mapping lemma should seem natural and elementary. This approach also serves to introduce students to fixed point techniques, which are powerful tools in many branches of mathematics, particularly differential equations and numerical analysis.

DEFINITION. Let $C \subset \mathbb{R}^n$. A mapping $T: C \to \mathbb{R}^n$ is a *contraction* if there exists a constant α with $0 \leqslant \alpha < 1$ such that

$$\|T(p) - T(q)\| \leqslant \alpha\|p - q\| \quad \text{for all } p, q \in C.$$

THEOREM (Contraction Mapping Lemma on $\mathbb{R}^n$). *Let* C *be a non-empty, closed subset of* $\mathbb{R}^n$, *and let* $T: C \to C$ *be a contraction. Then* T *has a unique fixed point in* C, *that is, a point* p *for which* $T(p) = p$.

Proof. Define $f: C \to \mathbb{R}$ by $f(x) = \|x - T(x)\|$. Note that a zero for f is a fixed point for T. It is easy to show that

$$|f(x) - f(y)| \leqslant (1 + \alpha)\|x - y\|,$$

and so f is continuous. If C is bounded, the max/min principle implies the existence of $p \in C$ such that $f(p)$ is a minimum. Then $f(p) \leqslant f(T(p)) \leqslant \alpha f(p)$. Since $f(p) \geqslant 0$ and $\alpha < 1$, we have $f(p) = 0$.

If C is not bounded, choose $q \in C$ and set

$$\tilde{C} = \{x \in C | f(x) \leqslant f(q)\}.$$

If $x \in \tilde{C}$, then

$$\begin{aligned}\|x - q\| &\leqslant \|x - T(x)\| + \|T(x) - T(q)\| + \|T(q) - q\| \\ &\leqslant 2f(q) + \alpha\|x - q\|.\end{aligned}$$

Hence

$$\|x - q\| \leqslant \frac{2f(q)}{1 - \alpha},$$

so $\tilde{C}$ is closed and bounded. From $f(T(p)) \leqslant \alpha f(p)$, it follows that T preserves $\tilde{C}$, and we may proceed as above.

Finally, if $p, q \in C$ are both fixed points of T, then

$$\|p - q\| = \|T(p) - T(q)\| \leqslant \alpha\|p - q\|,$$

and so $\|p - q\| = 0$ and the fixed point is unique.

In closing, we should point out that the contraction mapping lemma is the basis for several iterative numerical methods familiar to undergraduates, e.g., Newton's method. (For a good

survey of methods from this point of view, see [2].) In these applications one uses the fact that if x_0 is an arbitrary element of the set on which T is a contraction, then the sequence $x_0, x_1, x_2, \ldots$ defined by $x_{n+1} = T(x_n)$ converges geometrically to the fixed point. This property, which is a by-product of the usual proof of the contraction mapping lemma, follows easily from

$$\text{(i)} \qquad f(T(x)) \leqslant \alpha f(x).$$

If p is the fixed point, then

$$\|x - p\| \leqslant \|x - T(x)\| + \|T(x) - T(p)\| \leqslant f(x) + \alpha\|x - p\|,$$

and so

$$\text{(ii)} \qquad \|x - p\| \leqslant \frac{1}{1-\alpha} f(x).$$

Combining (i) and (ii) for the sequence $\{x_n\}$ yields

$$\|x_{n+1} - p\| \leqslant \frac{\alpha^n}{1-\alpha} f(x_0),$$

which shows that the convergence is at least geometric. Inequality (ii) gives a better estimate, however, and shows that the function f gives a convenient test as to when the iteration should stop. If one wishes x_n to be within ε of the fixed point, then the iteration should continue until $f(x_n) < \varepsilon(1 - \alpha)$.

References

1. S. Lang, Differential Manifolds, Addison-Wesley, Reading, MA, 1972.
2. C. H. Wagner, A generic approach to iterative methods, Math. Mag., 55 (1982) 259–273.

MAXIMIZING THE AREA OF A TRAPEZIUM

JOHN P. HOYT
1115 Marietta Avenue, #32, Lancaster, PA 17603

We define a trapezium as a convex quadrilateral with no parallel sides. Generally speaking a trapezium is uniquely determined by five independent conditions. For example, a trapezium is uniquely determined given the lengths of four consecutive sides and the fact that the sum of one pair of opposite angles equals the sum of the other pair. This last condition is the most well-known necessary and sufficient condition that a convex quadrilateral be inscribable in a circle [**1**, pp. 74 and 88]. Many interesting relationships exist between the parts of an inscribed quadrilateral. We shall have occasion to refer to Brahmagupta's Theorem [**2**, p. 86]: *If the sides of an inscribed quadrilateral are of lengths a, b, c, and d and if the semi-perimeter is denoted by s, the area of the quadrilateral is given by*

$$A = [(s-a)(s-b)(s-c)(s-d)]^{0.5}.$$

On the other hand a trapezium is not uniquely determined given the lengths of four consecutive sides and the fact that the sum of one pair of opposite sides equals the sum of the other pair. The last condition is the most well-known necessary and sufficient condition that a quadrilateral be circumscribable about a circle [**1**, p. 89]. There is usually an infinite number of such circles for each given set of sides. However, if in addition to the lengths of the four consecutive sides and the fact that the sums of the opposite pairs of sides are equal, we are given one of the segments on one of the sides made by the point of tangency, then the quadrilateral is uniquely determined and knowing one segment we know them all. If we let the segments made on side a by the point of tangency of a particular inscribed circle be p and q and the segments made on opposite side c by

the point of tangency of the same circle are t and u, then the area of the quadrilateral is given by

$$A = [s(pq[t+u] + tu[p+q])]^{0.5}.$$

I discovered this formula over forty years ago and a near-complete proof of it can be found in [**3**, p. 242, Q694].

The fact that there is an infinite number of circumscribable quadrilaterals with the same four consecutive sides suggests the following problem:

Find the maximum area of a convex quadrilateral whose consecutive sides have lengths of a, b, c, and d with $a + c = b + d$.

The formula for the area of a circumscribable quadrilateral given in the preceding paragraph is useful for a one variable solution of this problem. A two-variable solution follows.

If we let x be the length of the segment made by the point of tangency of the inscribed circle on the side of length a as in Fig. 1, we have $A = [sf(x)]^{0.5}$ with $s = a + c = b + d$ and

$$f(x) = x(a-x)c + (b-a+x)(d-x)a.$$

Since A is maximized when $f(x)$ is maximized, we differentiate $f(x)$ and find the critical value to be $x = ad/(b+d)$. Since $f(x)$ is a quadratic, it is apparent that the critical value of x yields a maximum value of $f(x)$ and A. Evaluating the maximum value is easily done if one observes that $x = ad/(b+d)$ implies $x/(a-x) = d/b$, and hence by symmetry we can see that the points of tangency of the inscribed circle divide each side in the ratio of the two adjoining sides. Then

$$a - x = ab/(b+d),\ b - a + x = bc/(b+d), \text{ and } d - x = dc/(b+d).$$

Substituting, we have $f(ad/[b+d]) = abcd/(b+d)$ and maximum $A = (abcd)^{0.5}$.

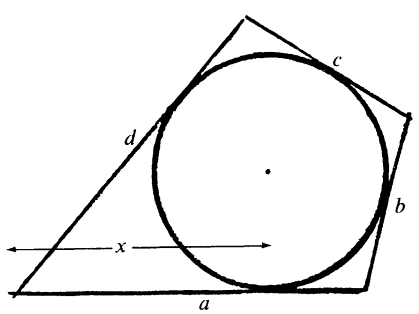


FIG. 1

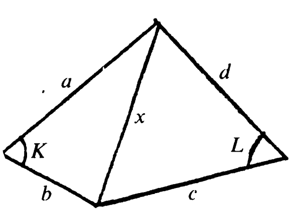


FIG. 2

Either the fact that the points of tangency of the inscribed circle divide the sides of the quadrilateral proportional to the adjoining sides, or the fact that the area of the quadrilateral is $[abcd]^{0.5}$ implies that the circumscribable quadrilateral with maximum area for four given sides is also inscribable [**1**, p. 322].

In case you think we didn't play fair in solving our problem with a little known theorem, read on.

We now suggest and solve the more general problem:

Find the maximum area of the convex quadrilateral whose consecutive sides have lengths of a, b, c, and d.

Referring to Fig. 2, we have

(1) $$A = 0.5(ab\sin K + cd\sin L)$$

and

$$a^2 + b^2 - 2ab\cos K = c^2 + d^2 - 2cd\cos L \tag{2}$$

(both $= x^2$).

Differentiating (1) and (2) with respect to K, we have

$$dA/dK = 0.5(ab\cos K + cd\cos L\,dL/dK) \tag{3}$$

and

$$ab\sin K = cd\sin L\,dL/dK. \tag{4}$$

For $dA/dK = 0$ subject to (4), we find $\operatorname{ctn} K = -\operatorname{ctn} L$ or angles K and L are supplementary since we are dealing with convex quadrilaterals. Differentiating (4) with respect to K, we find

$$d^2L/dK^2 = m\operatorname{ctn} K, \text{ where } m = (a^2b^2 + abcd)/c^2d^2.$$

Differentiating (3) with respect to K, we find that when $dA/dK = 0$, $d^2A/dK^2 = -0.5\,mcd\csc K$, which is negative for all possible values of K. Hence the area is maximized when the opposite pairs of angles are supplementary or the quadrilateral is inscribable. Using (2) we can see that

$$\cos K = (a^2 + b^2 - c^2 - d^2)/2(ab + cd)$$

when $dA/dK = 0$ and with some laborious calculations we can show that the maximum area is

$$[(s-a)(s-b)(s-c)(s-d)]^{0.5}, s = (a + b + c + d)/2.$$

Much more easily we can use Brahmagupta's Theorem to obtain the same result. It can be seen from the symmetry of the answer that the word "consecutive" can be removed from the statement of the problem.

The method used on the second problem is applicable to the first problem with the condition $a + c = b + d$ not used until after Brahmagupta's formula appears. Using $a + c = b + d$ in this formula will yield $A = [abcd]^{0.5}$ as before. $A = [abcd]^{0.5}$ is the not so well-known formula for the area of a trapezium which is both inscribable in and circumscribable about a circle.

Try these problems on your better students.

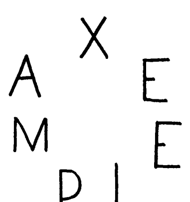
References

1. W. Chauvenet, Treatise on Elementary Geometry, J. B. Lippincott Co., Philadelphia, 1870.
2. F. Cajori, A History of Mathematics, Macmillan, New York, 1931.
3. J. P. Hoyt, Mathematics Magazine, Quickies, 4 (1984) pp. 239 and 242.

MORE REBUSES AGAIN

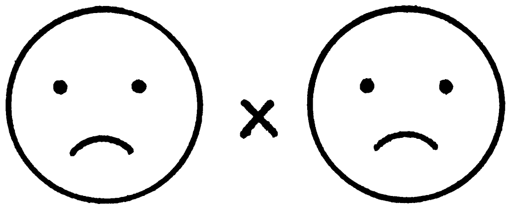
According to the American Heritage Dictionary, a *rebus* is a riddle composed of words or syllables depicted by symbols or pictures that suggest the words or syllables they represent. Can you decipher the ones below?

1.



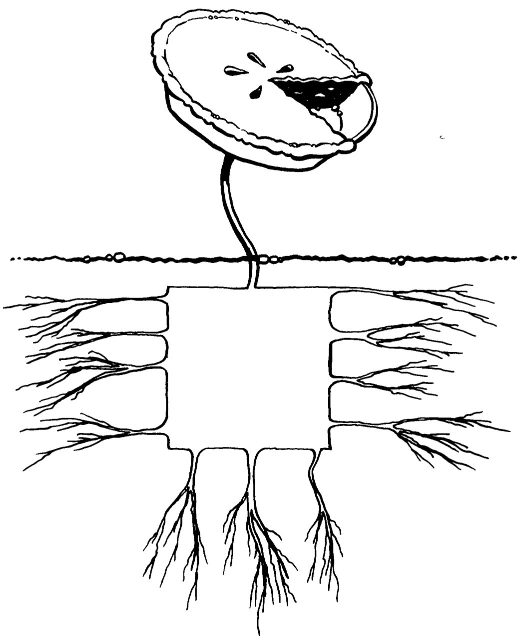
—Rochelle Leibowitz

2.



—Phyllis Lefton

3.



—Vincent P. Shielack, Jr.

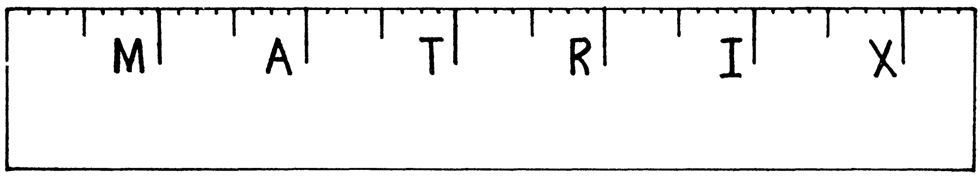
4.

die Ebene
el plano
le plan
mặt phẳng
samatal ti
ΤΟ ΕΠΙΠΕΔΟΝ
ПЛОСКОСТЬ
平 평
面 면
المسطح
సమతలము
המישור

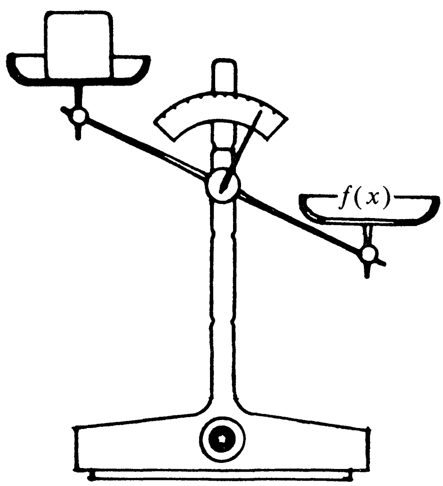
—W. Edwin Clark

5. M
A
TRIX

—R. B. Bapat

6. 

—R. B. Bapat

7. 

—Hosam M. Mahmoud

Answers on p. 75.

---

160. MISCELLANEA

To be freed from quantity is to be freed from exchange. Then the world becomes sacred again —not for sale. ... A corollary: people freed from the rule of quantity do not stop being shrewd or able to calculate, any more than they lose interest in science and mathematics. All they lose is the propensity to form technocratic delusions of grandeur—and the need to submit to technocratic authority.

—From *Against the State of Nuclear Terror*, Joel Kovel,
South End Press, Boston, 1983, p. 215.

The book of Wanda Szmielew answers for the first time all the questions thus raised. It presents the relationship between algebra and geometry at the earliest stage in the development of concepts, beginning with the weakest algebraic structures (which may be neither commutative nor associative) for which there are *reasonable* geometric equivalents. The key word for this relationship is "representation theorem". Indeed, there are a dozen representation theorems in this work, generalizing those of Hilbert and Tarski mentioned above.

In Szmielew's treatment, affine geometry is based on a quaternary relation among points called parallelity ("$ab \| cd$"), instead of being built up, as usual, on collinearity (a ternary relation "L", with "$L(abc)$" meaning "the points a, b, c are collinear"). These relations are mutually definable by:

$$ab \| cd \leftrightarrow \left(a = b \vee \forall p(L(abp) \to L(cdp)) \vee \forall p(L(abp) \to \neg L(cdp))\right)$$

and

$$L(abc) \leftrightarrow ab \| bc.$$

So, what is gained by this non-traditional approach to affine geometry? The resulting axiom system, for what the author calls "parallelity planes", is simpler, in the sense that parallelity planes admit an $\forall\exists$-axiom system, whereas affine geometry based on collinearity doesn't. (A $\forall\exists$-axiom system is one in which all the axioms are written in prenex form so that all universal quantifiers precede all existential quantifiers.)

The coordinatization of parallelity planes leads to a very weak algebraic structure called a "ternary field". If one adds the minor or major Desargues theorem then one gets quasi-fields or skew fields, respectively, as coordinate algebraic structures. If one adds instead Pappus' law the coordinate field becomes commutative.

In this way algebra and geometry develop in parallel. The odyssey of travelling to Euclidean geometry by way of affine geometry is the content of this marvellous book. Unfortunately, Ithaca is not reached. The author died before finishing the last chapter. Most gaps have been filled by the editor, M. Moszyńska, whose style is very close to that of the author.

This book leaves the reviewer with a feeling such as one experiences watching the throwing of the javelin at the Olympic Games. Every movement seems simple, natural, the way it has to be. One almost forgets that behind this apparent ease and grace there are years of hard work, of sweat and tears.

The present book is a life's work. As expected from an accomplished scholar, the language is simple and precise. The monograph is suitable for a beginning graduate course in the foundations of geometry.

The bibliography of the works of Wanda Szmielew at the end of the book is there to tell us that she praised above all in mathematics meaningfulness and beauty.

ANSWERS TO REBUSES ON PAGES 56–58

1. Counterexample.

2. Cross product.

3. $\Gamma(\frac{1}{2})$. (See p. 76.)

4. A group of translations of the plane.

5. Orthogonal matrix.

6. Scalar matrix.

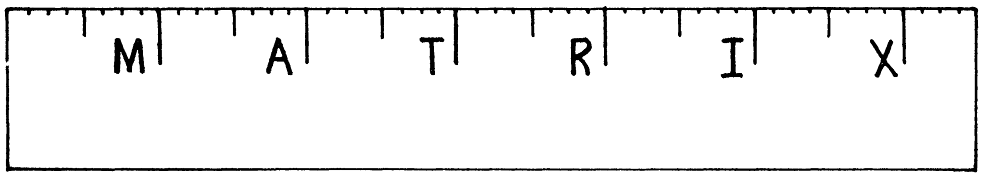
7. Heaviside function.

5.

M
A
TRIX

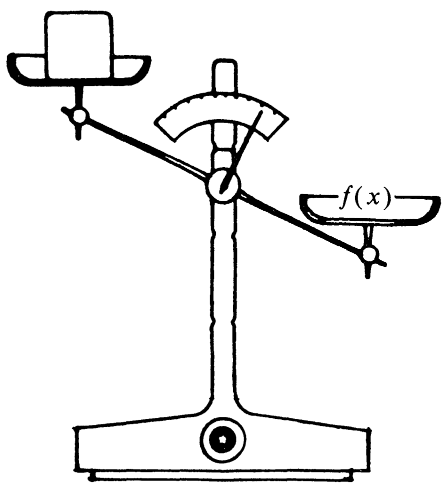
—R. B. Bapat

6.



—R. B. Bapat

7.



—Hosam M. Mahmoud

Answers on p. 75.

---

160. MISCELLANEA

To be freed from quantity is to be freed from exchange. Then the world becomes sacred again —not for sale. ... A corollary: people freed from the rule of quantity do not stop being shrewd or able to calculate, any more than they lose interest in science and mathematics. All they lose is the propensity to form technocratic delusions of grandeur—and the need to submit to technocratic authority.

—From *Against the State of Nuclear Terror*, Joel Kovel,
South End Press, Boston, 1983, p. 215.

PROBLEMS AND SOLUTIONS

Send all proposed problems, typed and in duplicate if possible, to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the address given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by May 31, 1986. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgement).

E 3123. *Proposed by R. M. Grassl and T. Porter, University of New Mexico.*

Prove that

$$\frac{(n^2)!}{\binom{n}{n}\binom{n+1}{n}\binom{n+2}{n}\cdots\binom{2n-1}{n}(n!)^n}$$

is an integer for $n \in \{1,2,3,\dots\}$.

E 3124. *Proposed by N. Gauthier, P. Rochon, and J. R. Gosselin, Royal Military College of Canada, Kingston, Ontario.*

Solve the following set of two coupled non-linear differential equations:

$$\frac{d}{dt}\frac{x}{(1-r^2)^{\frac{1}{2}}} = yf,$$

$$\frac{d}{dt}\frac{y}{(1-r^2)^{\frac{1}{2}}} = f - xf.$$

By definition, $r^2 = x^2 + y^2$; by assumption, $x = y = 0$ at $t = 0$ and $f = f(t)$ is t-integrable but otherwise arbitrary.

Proposers' remarks. The above set of equations arose in an effort to explain the physical phenomenon of radiation pressure (e.g., comets' tails point away from the sun) in elementary terms. Normal treatments follow Maxwell and use the electromagnetic stress tensor. Our treatment uses the relativistic equations of motion for a charge Q of rest mass m in an x-directed electromagnetic wave of wavelength λ. The equations of motion are of the form

$$\frac{d}{dt}\frac{mv_x}{(1-v^2/c^2)^{\frac{1}{2}}} = QB_z v_y,$$

and

$$\frac{d}{dt}\frac{mv_y}{(1-v^2/c^2)^{\frac{1}{2}}} = QE_y - QB_z v_x.$$

In these expressions, c is the speed of light in vacuo, v_x and v_y are the x- and the y- components of the velocity of the charge ($v^2 \equiv v_x^2 + v_y^2$), E_y and B_z are the assumed y- and z-directed electric and magnetic fields. If the forward spatial drift of the charge, in one cycle of the wave, is small compared to λ, then

$$E_y = B_z/c = E_0 \sin(2\pi ct/\lambda + \phi);$$

E_0 is a constant, and ϕ is a random initial phase (to be averaged out at the end of the calculation). These equations are to be solved for $\langle x \rangle$ and $\langle y \rangle$, the mean x- and y-drifts in one cycle, after averaging over ϕ has been done, to show that $\langle x \rangle \neq 0$ while $\langle y \rangle = 0$. (If the mean drift is comparable to λ, then E_y and B_z are as given but with ϕ now a function of x: $\phi = -2\pi x/\lambda +$ constant. We have not succeeded in solving this case exactly and would welcome suggestions.)

E 3125. *Proposed by I. J. Schoenberg, Madison, Wisconsin.*

Find two positive non-increasing sequences $\{a_n\}, \{b_n\}$ $(n = 1,2,\dots)$ such that $\sum_1^\infty a_n$ and $\sum_1^\infty b_n$ both diverge, while $\sum_1^\infty \min\{a_n, b_n\}$ converges.

E 3126. *Proposed by F. S. Cater, Portland State University.*

Let M be a 3 by 3 matrix with entries in a field F. Prove that M is similar over the field to precisely one of these three types:

$$\underset{\text{Type I}}{\begin{pmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{pmatrix}} \quad \underset{\text{Type II}}{\begin{pmatrix} b & 0 & 0 \\ 1 & c & 0 \\ 0 & 0 & c \end{pmatrix}} \quad \underset{\text{Type III}}{\begin{pmatrix} d & 1 & 0 \\ e & 0 & 1 \\ f & 0 & 0 \end{pmatrix}} \quad (a,b,c,d,e,f \in F).$$

E 3127. *Proposed by David Shelupsky, The City College of New York.*

(a) Show that, on any real interval $[a, b]$, the exponential function e^x is characterized, up to multiplication by an arbitrary positive constant, by the inequality

$$f(x) < (f(y) - f(x))/(y - x) < f(y),$$

where x and y are real, $a \leqslant x < y \leqslant b$.

(b)* Are the exponentials the only functions on an interval I that satisfy the inequality

$$\min\{f(x), f(y)\} \leqslant (f(y) - f(x))/(y - x) \leqslant \max\{f(x), f(y)\}$$

for $x \neq y$ in I?

E 3128. *Proposed by Florentin Smarandache, Lycée Sidi El Hassan Lyoussi, Sefrou, Morocco, Africa.*

If p is an odd prime, show that for all fixed $k \in \{1,2,\dots, p-2\}$, $\Sigma i_1 i_2 \cdots i_k$ is a multiple of p, where the summation is over all sequences $i_1, i_2, \dots, i_k$ such that $1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant p - 1$.

SOLUTIONS OF ELEMENTARY PROBLEMS

Isomorphisms of a Finite Lattice

E 2973 [1982, 698]. *Proposed by M. M. Konstantinov, Bulgaria.*

Let $(L, \leqslant)$ be a lattice, where L is a finite set and $\leqslant$ is an order relation. For $x, y \in L$ set $x \vee y = \sup(x, y)$ and $x \wedge y = \inf(x, y)$. Denote by P the set of the bijections $L \to L$. The bijection f is said to be increasing (decreasing) if $x \leqslant y$ implies $f(x) \leqslant f(y)$ $(f(x) \geqslant f(y))$. The bijection f is said to be monotone if it is either increasing or decreasing. The set P can be represented as a union $P = P^+ \cup P^- \cup P^0$ of disjoint components, where P^+, P^- and P^0 are the sets of increasing, decreasing, and nonmonotone bijections.

A. Let $f \in P$. Prove that the following three assertions are equivalent:

(i) $f(x \wedge y) = f(x) \wedge f(y)$;

(ii) $f(x \vee y) = f(x) \vee f(y)$;

(iii) $f \in P^+$.

B. Let $f \in P$. Prove that the following three assertions are equivalent:

(j) $f(x \wedge y) = f(x) \vee f(y)$;

(jj) $f(x \vee y) = f(x) \wedge f(y)$;

(jjj) $f \in P^-$.

Solution by Edwin L. Marsden, Norwich University, Northfield, VT.

LEMMA. *Let f: $L \to L$ be an increasing bijection on a finite partially ordered set. Then f is an order isomorphism, i.e., f has an inverse that also preserves order.*

Proof. We may assume that $f \neq 1$, the identity function. The function f preserves order, so f^n preserves order for all $n \in Z^+$. Since L is finite, for some positive integers $p > q$, $f^p = f^q$, and thus $f^{p-q} = 1$. Hence f^{p-q-1} is an inverse of f that preserves order.

A. Let $f \in P$. These are equivalent:

(i) $f(x \wedge y) = f(x) \wedge f(y)$;

(ii) $f(x \vee y) = f(x) \vee f(y)$;

(iii) $f \in P^+$.

Proof. (i) $\Rightarrow$ (iii). For $x \leqslant y$ then $f(x) = f(x \wedge y) = f(x) \wedge f(y)$, so $f(x) \leqslant f(y)$.

(ii) $\Rightarrow$ (iii) is equally direct.

(iii) $\Rightarrow$ (ii). By the lemma f is an order isomorphism and $f^{-1} = f^m$ for some $m \in Z^+$. We wish to show that $f(x \vee y) = f(x) \vee f(y)$. By hypothesis $f(x)$, $f(y) \leqslant f(x \vee y)$. Suppose that $f(x)$, $f(y) \leqslant u \in L$. Then

$$x = f^m f(x) \leqslant f^m(u) \quad \text{and} \quad y \leqslant f^m(u),$$

so $x \vee y \leqslant f^m(u)$. Therefore

$$f(x \vee y) \leqslant ff^m(u) = u.$$

Thus $f(x) \vee f(y) = f(x \vee y)$.

(iii) $\Rightarrow$ (i) follows similarly.

B. For $f \in P$, these are equivalent:

(j) $f(x \wedge y) = f(x) \vee f(y)$;

(jj) $f(x \vee y) = f(x) \wedge f(y)$;

(jjj) $f \in P^-$.

Proof. (j) ⇒ (jjj) and (jj) ⇒ (jjj) are simple.

(jjj) ⇒ (j). Notice that $f \in P^-$ implies that $f^2 \in P^+$. Since $f \in P^-$, then $f(x)$, $f(y) \leqslant f(x \wedge y)$. Suppose that $f(x)$, $f(y) \leqslant u \in L$. It follows that $f(u) \leqslant f^2(x)$, $f^2(y)$. By part A,

$$f^2(x) \wedge f^2(y) = f^2(x \wedge y).$$

Thus

$$f(u) \leqslant f^2(x \wedge y) \quad \text{and} \quad f^2 f(x \wedge y) = f^3(x \wedge y) \leqslant f^2(u).$$

Apply the lemma to f^2 to get $f(x \wedge y) \leqslant u$. Hence $f(x) \vee f(y) = f(x \wedge y)$.

(jjj) ⇒ (jj) is proved similarly.

Also solved by A. Bager (Denmark), Chico Problem Group, U. Faigle (West Germany), F. Gerrish (England), M. Hébert (Canada), O. P. Lossers (The Netherlands), O. Matouš (Czechoslovakia), J.-M. Monier (France), R. Patenaude, Rand Afrikaans University Problem Solving Group (South Africa), R. Richter and H. Shank (Canada), J. H. Riley, J. Schmid (Switzerland), A. Smuckler (Israel), G. P. Wene, M. Woltermann, Pei Yuan Wu (China), and the proposer. Several readers noted that P^+ and P^- aren't (quite) disjoint and that the finiteness condition on L is necessary.

Evaluation of an Infinite Series

E 2976 [1982, 756]. *Proposed by Lee Whitt, D. H. Wagner Associates, Hampton, Virginia.*

Let N_0 be a given nonnegative integer and p be a real number such that $0 < p < 2$. Give a closed form expression for the summation

$$\sum_{L=N_0}^{\infty} \binom{L}{N_0} p^{N_0}(1-p)^{L-N_0}$$

where we define $0^0 = 1$ for $p = 1$ and $L = N_0$. Note that the solution is independent of N_0.

Solution by Howard Wiener (student), Vanderbilt University. For $0 < p < 2$, give a closed form expression for

$$\sum_{L=N_0}^{\infty} \binom{L}{N_0} p^{N_0}(1-p)^{L-N_0},$$

where we take $0^0 = 1$ for $p = 1$ and $L = N_0$.

The value of this sum is $1/p$. Before proceeding to show this, we note that $0 < p < 2 \Rightarrow |1 - p| < 1$. From this fact, use of the ratio test shows that the series converges for N_0 any nonnegative integer. We now define $q = 1 - p$ for convenience, and also:

$$f(N) = \sum_{L=N}^{\infty} \binom{L}{N} p^N q^{L-N}.$$

Two solutions are now given.

SOLUTION I. We note that $f(0)$ is a simple geometric series:

$$f(0) = \sum_{L=0}^{\infty} q^L = \frac{1}{1-q} = \frac{1}{p}.$$

Thus, we need only show that for N any nonnegative integer, $f(N) = f(0)$. This reduces (via an induction argument) to showing $f(N+1) = f(N)$. We note that:

$$f(N+1) = \sum_{L=N+1}^{\infty} \binom{L}{N+1} p^{N+1} q^{L-N-1} = \sum_{L=N}^{\infty} \binom{L+1}{N+1} p^{N+1} q^{L-N}$$

and recall that:

$$\binom{L+1}{N+1} = \binom{L}{N} + \binom{L}{N+1},$$

where it will be understood that a binomial coefficient is zero if the bottom number exceeds the top. Using these facts, we have:

$$\begin{aligned} f(N+1) - f(N) &= \sum_{L=N}^{\infty} \binom{L}{N} [p^{N+1} q^{L-N} - p^N q^{L-N}] + \binom{L}{N+1} p^{N+1} q^{L-N} \\ &= (p-1) \sum_{L=N}^{\infty} \binom{L}{N} p^N q^{L-N} + \sum_{L=N}^{\infty} \binom{L}{N+1} p^{N+1} q^{L-N} \\ &= -qf(N) + q \sum_{L=N+1}^{\infty} \binom{L}{N+1} p^{N+1} q^{L-N-1} \\ &= -qf(N) + qf(N+1) \end{aligned}$$

so that

$$f(N+1) - f(N) = q[f(N+1) - f(N)] \Rightarrow q = 1 \quad \text{or} \quad f(N+1) = f(N).$$

Since $p > 0$, the case $q = 1$ is impossible, so we are finished.

SOLUTION II. This is a somewhat more direct way to proceed. In the following, the operator D will denote differentiation with respect to q. For $N \leqslant L$, we have:

$$D^N q^L = \frac{L!}{(L-N)!} q^{L-N} = N! \binom{L}{N} q^{L-N}$$

so that an arbitrary term in the series $f(N)$ becomes:

$$\binom{L}{N} p^N q^{L-N} = (1-q)^N \binom{L}{N} q^{L-N} = \frac{(1-q)^N}{N!} D^N q^L$$

and thus $f(N)$ becomes:

$$f(N) = \frac{(1-q)^N}{N!} \sum_{L=N}^{\infty} D^N q^L.$$

To evaluate the series above, we need only note that since $|q| < 1$ the series $\sum_{L=0}^{\infty} q^L$ is a power series defining an analytic function of q, and is a geometric series converging to $1/(1-q)$. This gives us:

$$\sum_{L=N}^{\infty} D^N q^L = \sum_{L=0}^{\infty} D^N q^L = D^N \sum_{L=0}^{\infty} q^L = D^N \frac{1}{1-q} = \frac{N!}{(1-q)^{N+1}}.$$

Our final result is:

$$f(N) = \frac{(1-q)^N}{N!} \frac{N!}{(1-q)^{N+1}} = \frac{1}{1-q} = \frac{1}{p}$$

as we wished to show.

Also solved by 61 readers and the proposer. Garrett Sylvester cited the *Handbook of Mathematical Functions* (Abramowitz and Stegun, editors), page 822.

Lattice Points and the Taxicab Metric

E 2989 [1983, 212]. *Proposed by M. Goldberg, University of Waterloo, and S. C. Locke, Florida Atlantic University.*

Let $A = (a, a')$ and $B = (b, b')$ be lattice points (points in the plane with integral coordinates), and let $d(A, B)$ denote $|a - b| + |a' - b'|$. Let S be the set of points at d-distance at most k from the origin. Calculate

$$f(k) = \sum_{A,B \in S} d(A, B).$$

Solution by William A. Newcomb, Lawrence Livermore National Laboratory.

$$f(k) = \frac{2}{15}k(k+1)(2k+1)(7k^2+7k+6),$$

if the sum is understood to be over all *unordered* pairs of points A, B in S.

Proof. Horizontal and vertical segments contribute equally to the sum. Therefore,

$$f(k) = \sum_{-k \leqslant a < b \leqslant k} 2N_a N_b (b - a),$$

where N_x, for any integer x in the specified range, denotes the number of points in S with the horizontal coordinate x, namely, $2k + 1 - 2|x|$. We break up the sum as follows into three partial sums:

$$\sum_{0<a<b} + \sum_{a<b<0} = 2\sum_{0<a<b} 2N_a N_b (b-a),$$

$$\sum_{a=0<b} + \sum_{a<b=0} = 2\sum_{b>0} 2bN_0 N_b,$$

$$\sum_{a<0<b} 2N_a N_b (b-a) = \sum_{0<a,b} 2N_a N_b (b+a)$$

$$= 2\sum_{0<a<b} 2N_a N_b (b+a) + \sum_{b>0} 4bN_b^2.$$

The grand total is

$$\sum_{1}^{k} 4bN_b(N_0 + 2N_1 + \cdots + 2N_{b-1} + N_b),$$

wherein $N_b = 2k + 1 - 2b, N_0 + 2N_1 + \cdots = 2b(2k + 1 - b)$. The result, then, is

$$f(k) = 8(2k+1)^2 S_2 - 24(2k+1)S_3 + 16S_4,$$

where $S_n = 1^n + 2^n + \cdots + k^n$. Finally, by using known formulas for the power sums, one reduces this to the polynomial given above.

Also solved by J. Dou (Spain), M. Golomb, D. Hamlin, H. Honkasalo (Finland), J. B. M. Melissen (The Netherlands), W. Mixon, R. B. Nelsen, T. T. Nguyen, G. Sylvester, University of Arizona Problem Solving Group, and the proposers.

ANSWER TO PHOTO ON PAGE 24

Marston Morse, famous for, among other things, his book *The Calculus of Variations in the Large*.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by May 31, 1986. The solver's full post-office address should be on each sheet.

6507. *Proposed by David Callan, University of Bridgeport, Connecticut.*

Let P be an $r \times r$ stochastic matrix. It is known that P^n is Cesàro summable to a matrix A, i.e.,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} P^k = A, \quad \text{and} \quad Z = (I - (P - A))^{-1}$$

exists.

Find $\min\{tr(Z)\}$ and when it occurs.

6508. *Proposed by D. J. Newman, Temple University.*

The Famous "Nullstellensatz" can be stated as follows: Let $P_1, P_2, \ldots, P_k$ be any polynomials in variables $x_1, x_2, \ldots, x_n$. If the P_i have no common zero, then there are polynomials Q_i such that $\sum_{i=1}^{k} Q_i P_i$ has no zero at all.

This is usually proved when the underlying field is algebraically closed. Prove that it holds over any field.

SOLUTIONS OF ADVANCED PROBLEMS

Transformation of an Integral

6463 [1984, 440]. *Proposed by Jose M. Bayod, Santander, Spain.*

Let $S \subset R^n$ be a Lebesgue-measurable set with finite measure $\mu(S)$. Assume $f\colon S \to R$ is a real function that can be decomposed in the following way: $f = h \cdot g$, with g real, measurable and bounded (call $a = \inf g(S)$ and $b = \sup g(S)$), and h absolutely continuous on $[a, b]$. Then prove that f is integrable over S, and

$$\int_S f = \mu(S) h(b) - \int_a^b h'(t) \mu(g^{-1}[a, t])\, dt.$$

Solution by Rafael Obaya Garcia, Universidad de Valladolid, Spain. Since f is bounded and measurable on S and $\mu(S) < \infty$, it follows that f is integrable over S.

The function $G\colon [a, b] \times S \to R$ defined by

$$G(t, x) = h'(t) \chi_{g^{-1}([a,t])}(x)$$

is measurable on $[a, b] \times S$, and

$$\int_S dt \int_a^b |G(t, x)|\, dx \leqslant \mu(S) \int_a^b |h'(t)|\, dt < \infty.$$

This proves that G is integrable over $[a, b] \times S$ and so, by Fubini's Theorem,

$$\int_S dx \int_a^b G(t, x)\, dt = \int_a^b dt \int_S G(t, x)\, dx.$$

The left-hand side of this equation equals

$$\int_S dx \int_{g(x)}^{b} h'(t)\, dt = \int_S (h(b) - h(g(x)))\, dx = \mu(S) h(b) - \int_S f(x)\, dx,$$

and the right-hand side equals

$$\int_a^b h'(t)\,dt \int_S \chi_{g^{-1}([a,t])}(x)\,dx = \int_a^b h'(t)\mu(g^{-1}[a,t])\,dt.$$

The desired result follows.

Also solved by K. B. Athreya, P. J. Fitzsimmons, B. L. Granovsky (Israel), John Morrison, William A. Newcomb, Victor Pambuccian (Romania), Klaus Schürger (Federal Republic of Germany), Rae Michael Shortt, J. Vukmirović (Yugoslavia), and the proposer.

Intervals of Monotonicity of Polynomials

6465 [1984, 440–441]. *Proposed by F. S. Cater, Portland State University.*

Prove that for each integer $n \geqslant 3$, there is a largest number h_n such that there exists a polynomial p of degree at most n which increases on the interval $(-h_n, h_n)$ and satisfies $p(1) \leqslant p(-1)$. Prove that $0 < h_{2k-1} = h_{2k} < 1$ for $k \geqslant 2$, and that $\lim_{n\to\infty} h_n = 1$. (Compare with Problem 6449.)

Solution by Miroslav D. Ašić, University of Belgrade, Yugoslavia. Let k be an integer larger than 1. By the Legendre-Gauss quadrature formula, if $f(x)$ is any polynomial of degree at most $2k-1$, then

$$(*) \qquad \int_{-1}^{1} f(x)\,dx = \sum_{i=1}^{k} H_i f(x_i),$$

where $x_1, x_2, \ldots, x_k$ are the zeros of the kth Legendre polynomial with $x_1 < x_2 < \cdots < x_k$ and the coefficients H_i are all positive. We shall show that $h_{2k-1} = h_{2k} = x_k$. Note that $0 < x_k = -x_1 < 1$. Let

$$p'(x) = -(x-x_1)(x-x_2)^2 \cdots (x-x_{k-1})^2(x-x_k).$$

Clearly $p'(x) \geqslant 0$ for $x_1 < x < x_k$ so that p increases on (x_1, x_k). Since $\deg p = 2k-1$ and $p'(x_i) = 0$ for $i = 1,2,\ldots,k$, it follows from $(*)$ that

$$p(1) - p(-1) = \int_{-1}^{1} p'(x)\,dx = 0$$

and hence that $x_k \leqslant h_{2k-1}$.

On the other hand, let $x_k < r < 1$ and suppose that there is a polynomial q of degree at most $2k$ which increases on $(-r, r)$ and satisfies $q(1) \leqslant q(-1)$. Applying $(*)$ to q', we conclude from the inequalities $q'(x_i) \geqslant 0$ and $H_i > 0$ for $i = 1,2,\ldots,k$ that

$$q'(x_i) = 0 \quad \text{for } i = 1,2,\ldots,k.$$

Moreover, since $q'(x) \geqslant 0$ on $(-r, r)$, we also have

$$q''(x_i) = 0 \quad \text{for } i = 1,2,\ldots,k.$$

But $q'(x_i) = q''(x_i) = 0$ for $i = 1,2,\ldots,k$ implies that q is constant since $\deg q \leqslant 2k$. Therefore $h_{2k} \leqslant x_k$. Since $h_{2k-1} \leqslant h_{2k}$, it follows that

$$0 < h_{2k-1} = h_{2k} = x_k < 1.$$

It remains to prove that $\lim_{n\to\infty} h_n = 1$. Let $q'(x) = (2k+1)^{-1} - x^{2k}$. Then $q(1) = q(-1)$ and q increases on $(-a_k, a_k)$ where $a_k = (2k+1)^{-1/2k}$. Hence

$$1 > h_{2k-1} = h_{2k} \geqslant a_k \to 1 \text{ as } k \to \infty,$$

and so $h_n \to 1$ as $n \to \infty$.

Also solved by Ulrich Everling (West Germany), N. J. Lord (England), The University of South Alabama Problem Group, J. Vukmirović (Yugoslavia), and the proposer.

A Vacuous Property

6469 [1984, 518]. *Proposed by A. Wilansky, Lehigh University.*

1. Suppose that a countable set E of real sequences has the property that $E^\beta = l = \{x: \Sigma|x_n| < \infty\}$. Show that E has a finite subset with the same property. (Notation: $E^\beta = \{x: \Sigma x_i y_i$ converges $\forall y \in E\}$.)
2. Does such a set E exist?

Solution by Erich Badertscher, Mathematical Institute, Bern, Switzerland. Suppose that E is a countable set of real sequences and that $l \subseteq E^\beta$. We shall show that $l \neq E^\beta$.

Since $l \subseteq E^\beta$, it follows that $E \subseteq m$, the set of bounded real sequences. Next, we have that $E = \{y^{(n)}: n \in \mathbb{N}\}$. We define a sequence of infinite subsets $\{I_n\}$ of $\mathbb{N}$ as follows: Set $I_0 = \mathbb{N}$ and suppose I_k has been defined. Let $i_k = \min(I_k)$, and choose I_{k+1} to be an infinite subset of $I_k \setminus \{i_k\}$ such that $y^{(k+1)}$ is monotonic on I_{k+1}. Set $I = \{i_1, i_2, \dots\}$. Then each $y \in E$ is bounded and eventually monotone on I. Now define the sequence $x = \{x_n\}$ by setting

$$x_n = \begin{cases} (-1)^k/k & \text{if } n = i_k, \quad k = 1, 2, \dots, \\ 0 & \text{otherwise.} \end{cases}$$

Then $x \notin l$, but $x \in E^\beta$ since, for $y = \{y_n\} \in E$,

$$\sum_{n=1}^{\infty} x_n y_n = \sum_{k=1}^{\infty} \frac{(-1)^k}{k} y_{i_k}$$

is convergent by Abel's test.

Thus the answer to part 2 is "no", and part 1 is vacuously true.

Also solved by Fuensanta Andreu (Spain), David Hecker, and the proposer.

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER

Selected Papers of Loo-Keng Hua. Edited by H. Halberstam. Springer-Verlag, New York, 1983. xiv + 889 pp.

PAUL T. BATEMAN
Department of Mathematics, University of Illinois, Urbana, IL 61801

The career of Loo-keng Hua is one of the most interesting phenomena of recent mathematical history. Starting from humble beginnings and acquiring only nine years of formal schooling, he successfully made the transition from the role of a self-taught mathematical wunderkind to that of a fully accomplished and versatile creative mathematician of very high order. In addition he has been a leader in promoting the broadest possible application of mathematical tools in practical situations and in stressing the importance of good mathematical pedagogy. He has exerted great influence through his direction of research both in the United States and in mainland China. Finally he has been one of the foremost statesmen of science in the People's Republic of China.

Hua was born in 1909 and, aside from a stay in England from 1936 to 1938 and a stay in the U.S.A. from 1946 to 1950, has spent his entire life in China. His reappearance on the international

mathematical scene in 1978, after over a quarter of a century of minimal contact with Western mathematicians, was a very happy occasion.

Hua clearly wins the prize for being the contemporary mathematician known to the largest number of inhabitants of the earth. In China he is so widely revered that a television mini-series has been made of his early life. No present-day Western mathematician comes close to this level of recognition among the general public.

It is natural to compare Hua with the celebrated Indian mathematician Srinivasa Ramanujan, of an earlier generation, since their careers have striking similarities and at the same time clearcut differences. Both were largely self-taught, both profited from a period of research in England under the aegis of G. H. Hardy, each played a certain role in bridging the gap between East and West and in bringing his native land into the world of mathematical research, and each eventually achieved the status of a scientific folk-hero in his own country in the same way that Einstein did in the United States. On the other hand, there are stark contrasts between the two men. First, Ramanujan did not make the full transition from the role of self-taught genius to that of a sophisticated and trained mathematician, but remained a mathematical primitive to some extent and even retained a certain enigmatic character. Hua somehow became a mainstream mathematician fairly early in his mathematical career. Second, the contact with G. H. Hardy was much more direct and decisive in the case of Ramanujan; e.g., many of Ramanujan's papers after he came to England in 1914 went to the printer in Hardy's handwriting. Although Hua profited greatly from his stay in England, his mathematical contacts were apparently not as sharply focussed. Finally, while Ramanujan apparently did not have a great talent for survival and adaptability to differing living conditions, Hua has been in some sense a survivor and clearly has been able to adjust to differing academic, political, and dietary conditions.

A mathematician has to be judged by his research accomplishments and not by the number of university degrees earned. In Hua's case there are many of the former and none of the latter. Hua's research extends into the fields of number theory, algebra, geometry of matrices, classical groups, several complex variables, harmonic analysis, and applied mathematics.

Hua is probably best known for his work in number theory, particularly on Waring's problem. Waring's problem concerns itself with representing positive integers as sums of a bounded number of kth powers for a particular value of k. Restrictions on the integers or on the kth powers may be imposed and generalizations to other integral-valued polynomials are possible. A sample result of Hua's is his theorem that every sufficiently large odd positive integer is expressible as a sum of nine cubes of prime numbers. For experts in analytic number theory his most significant result is his theorem that if $q, a_1, a_2, \dots a_k$ are positive integers without common factor, and ϵ is a positive number, then

$$\left|\sum_{x=1}^{q} e^{2\pi i(a_1x+a_2x^2+\cdots+a_kx^k)/q}\right| \leqslant cq^{1-1/k+\epsilon},$$

where c is a constant depending only on k and ϵ. Another important result is what is known as Hua's Lemma or Hua's Inequality, namely

$$(*)\qquad \int_0^1\left|\sum_{m=1}^{N} e^{2\pi i\alpha m^k}\right|^{2^k} d\alpha \leqslant cN^{2^k-k+\epsilon},$$

where again c is a constant depending on the positive integer k and the positive number ϵ. The inequality $(*)$ makes possible a fairly short proof that every sufficiently large positive integer is expressible as a sum of 2^k+1 or fewer kth powers of positive integers. (This short proof is presented, for example, as Chapter 2 of R. C. Vaughan's book, *The Hardy-Littlewood Method*, Cambridge University Press, 1981.)

As a sample of Hua's work outside of number theory, we state two simple results of Hua on the basic properties of a skew-field or division ring F.

(1). *If σ is a one-to-one mapping of F onto itself such that $1^\sigma = 1$ and*

$$(a+b)^\sigma = a^\sigma + b^\sigma, \qquad (aba)^\sigma = a^\sigma b^\sigma a^\sigma,$$

for all a, b in F, then either $(ab)^\sigma = a^\sigma b^\sigma$ for all a, b in F or $(ab)^\sigma = b^\sigma a^\sigma$ for all a, b in F.

(2). *If $ab \neq ba$, then*

$$a = \left\{ b^{-1} - (a-1)^{-1} b^{-1} (a-1) \right\} \left\{ a^{-1} b^{-1} a - (a-1)^{-1} b^{-1} (a-1) \right\}^{-1},$$

an identity which makes immediate the so-called Brauer-Cartan-Hua theorem that *every proper normal sub-skew-field of a skew-field is contained in its center.*

These two results are "not so deep as a well, nor so wide as a church-door," but, like Mercutio's wound, they will serve.

Most of the research for which Hua is best known in this country was done before his return to China in 1950. One could speculate that he might have accomplished more in his personal research program if he had remained in the West; however, if he had done so, he would have been unable to play the central role that he did in the development of mathematics and its applications in China during the last three decades. In any case, Hua's total of scientific achievements is of a level justifying election to any learned society or academy.

The volume under review contains about one-third of Hua's papers. As is common nowadays, the composition was made by photographic reproduction from the original journals, resulting in a wide range of typographical quality. The choice of papers is a very reasonable one, but it was undoutedly influenced by the fact that many of Hua's papers had to be excluded out-of-hand because they were published in journals with poor typography.

The Selected Papers are a fitting monument to one of the great gentlemen of twentieth century mathematics.

Computability and Unsolvability. By Martin Davis. Dover Publications, New York, 1982. xxv + 248 pp. $6.50 (paperback).

MICHAEL STOB
Department of Mathematics and Computer Science, Calvin College, Grand Rapids, MI 49506

The open problem of theoretical computer science that gets the most press is certainly the $P = NP$ problem. To see the issues involved, consider the knapsack problem. (Theoretical computer scientists love cute names for problems. The Dining Philosophers and the Byzantine Generals are two recent examples.) We have a knapsack of capacity c units, c an integer. We also have n objects of integer sizes $a_1, \dots, a_n$ units respectively. The question is whether there is a subset of the n objects that will fit in the knapsack exactly; i.e., the sum of the sizes of the objects in the subset is exactly c. Of course there is a simple procedure that will answer our question—try all subsets! The defect of this solution is that it is slow. Precisely, the time it takes to run this algorithm grows exponentially in the number of objects. Surely, for a computationally feasible solution we would like an algorithm whose running time grows only polynomially in the number of inputs. (Actually we would hope that the polynomial is also of small degree with small constants but this is a start.) Were there such an algorithm, we would say that the knapsack problem is in the class of problems P. Whether it is in P is still open.

Suppose now however that we are good guessers and that the answer to our question is yes. Then there is an "algorithm" that quickly will verify that the answer is yes: we choose the correct subset and merely add up the sizes to verify that this is a solution. This algorithm gives yes answers in time linear in the number of inputs—we use linear time to write down the guess, linear

time to add up the sizes in the guessed subset. The "algorithm" shows that the knapsack question is in the class of problems *NP*—problems for which the answer yes can be proved in time polynomial in the length of the input by an algorithm which is allowed to make guesses. Notice that there is an asymmetry in this definition. Only the yes answers need be proved in polynomial time. If the complementary question to the knapsack question is also in *NP*, we say that the knapsack question is in *CO-NP*. Whether the knapsack question is in *CO-NP* is also unknown. Although the class *NP* intuitively seems to be much larger than the class *P*, whether $P = NP$ is still an open problem.

Surely to attack the $P = NP$ problem, and indeed even to formulate it exactly, a formal model is needed in which the informal notions of computation, algorithm, and complexity of computation are given precise meanings. The first such formal model arose out of mathematical logic and was used by Gödel to prove his celebrated Incompleteness Theorem. This theorem says roughly that there is no algorithm which, when given a purported theorem of arithmetic, will determine whether or not it has a proof from a given set of axioms. Due to the book *Gödel, Escher, Bach*, by Hofstadter, this theorem and its method of proof are now widely known even among undergraduates. Since Gödel, a number of alternative models of computation have been posed, the most influential of which has been the model of Alan Turing known as the Turing machine. All of these models are equivalent in one sense: a computable function in one model is computable in all other models. This is evidence to support the claim that the class of functions computable by an algorithm is precisely that defined by Gödel. (Of course this claim, which is known as Church's Thesis, cannot be proved since it equates an informal notion with a formal one.) Since Gödel's theorem, the theory of computability has not only developed as a rich mathematical theory, but also has been applied to establish both computability and non-computability results in a wide variety of areas.

The book under review is twenty-five years old. It is a re-release by Dover of a classic text on computability theory. Dover does the mathematical community a great service with this series which contains some masterpieces of mathematical exposition at very low prices. In this book, we find a clear and careful presentation of the state of the art in computability theory as of 1955 or so. The two models of computation of Gödel and Turing are presented and their equivalence proved. The basic theorems of the pure theory, including those on relative computability and the Post hierarchy of unsolvable problems, are presented. The applications described include Gödel's theorem (a weak form) and Post's proof that the word problem for semigroups is unsolvable. (The word problem for a class of algebraic structures is the problem of determining, given a finite set of equations in some generators of a structure in that class, whether some further equation $x = y$ is a consequence of that set of equations.) This account is extremely well-written and it surely must have served to introduce computability theory to a much wider audience than had previously considered the subject.

But the book is twenty-five years old and it has the obvious defects for an old book on a subject that was still relatively new when it was published. For instance, you won't find anything in this book about the applications of computability theory to computer science. The definitions and basic results of computational complexity theory, including the statement of the $P = NP$ problem, all were born in the sixties and seventies. Since the audience for which this book might be used as a text contains many prospective computer scientists, this is an important topic for such a text. I think a mathematician interested in recursion theory, particularly as applied to theoretical computer science, would be much better served by say Machtey and Young, *An Introduction to the General Theory of Algorithms*. Also, the book is out of date with respect to the results of the pure theory of computability. In many cases this is no problem, but in other instances even more elementary texts provide stronger theorems with proofs that are easier to understand. This is due in part to the fact that Davis' notation is not always what is now the standard; the relatively formal, non-intuitive style of argument in his book has been abandoned for more intuitive arguments; modern notation and terminology reinforces that intuition. The

standard source for the pure theory of compatibility is still the book by Rogers, *Theory of Recursive Functions and Effective Computability*. There are many more "modern" treatments suitable as textbooks for undergraduates. Finally, better accounts for the applications of computability theory exist. In fact, Davis himself has written a beautiful survey of what is known about undecidable problems for the *Handbook of Mathematic Logic*.

But having criticized an old text for being old, let me say that there is one feature of this book that makes it well worth the $6.50 anyway. In an appendix, Dover has reproduced Davis' beautiful paper, *Hilbert's Tenth Problem Is Unsolvable*, which originally appeared in this MONTHLY. The paper is a completely self-contained exposition of the proof that there is no algorithm for determining whether an arbitrary Diophantine polynomial equation with integer coefficients has an integer solution. It is a masterfully written article and, given to an undergraduate, is sure to awaken interest in computability theory. I know, I read it as an undergraduate and have been in love with recursion theory every since.

Differential Geometry in the Large. By Heinz Hopf. Lecture Notes in Mathematics, Vol. 1000. Springer-Verlag, Heidelberg-Berlin, 1983. vi + 184 pp. $9.50.

ROBERT OSSERMAN
Department of Mathematics, Stanford University, Stanford, CA 94305

Spending a year as a student in Zürich, I quickly learned that the optimal use of class time was to attend all courses of Heinz Hopf on whatever subject. The depth of his insight and the clarity of his presentation made him one of the great mathematical lecturers of his generation. Later on, I had the good fortune to be at Stanford when Hopf was a visiting Professor and to hear him lecture on one of the subjects to which he had himself contributed so profoundly: global differential geometry. Those lectures were written up and distributed in several *samizdat* versions over the years, partly in response to a steady stream of requests for copies. Now Springer-Verlag has done a great service by publishing the notes, together with an earlier set of lectures given at New York University, as Volume 1000 in the series "Lecture Notes in Mathematics." The version they published was the original set of notes prepared by John Gray in 1956; the NYU notes, by Peter Lax, date from 1946. Thus, this volume may be said to qualify as an instant classic.

Roughly speaking, Part I (the NYU notes) focuses on the piecewise-linear approach, and Part II (the Stanford notes) on the differentiable category. Thus, Part I has an extended discussion of polyhedra, including standard topics such as the Euler-Poincaré formula, and much more unusual ones, like a proof of Cauchy's theorem that convex polyhedra are rigid, and the solution by Max Dehn of Hilbert's third problem. That problem, as Hilbert notes, goes back to Gauss, who pointed out that for some elementary theorems in solid geometry there were no known elementary proofs. For example, to prove that two triangular pyramids with the same height and same area of base had the same volume seemed to require calculus (or arguments by exhaustion). This is in contrast to the two-dimensional case, where two polygons have the same area if and only if they can be decomposed into a finite number of pairwise congruent triangles. In three dimensions, a regular tetrahedron and a cube of the same volume *cannot* be decomposed into a finite number of pairwise congruent polyhedra, which settles the problem of Gauss/Hilbert.

Among the many items of piecewise-linear lore in this volume that are far from well known, consider the following question: What is the 3-dimensional analog of the fundamental formula of plane geometry, that the sum of angles in any triangle is π? Answer: the formula of de Gua (1783): For any tetrahedron, the sum of the dihedral angles minus half the sum of the solid angles equals 2π.

We may note in passing that Springer-Verlag has given the title "Differential Geometry in the

Large" of the Stanford notes to the volume as a whole. However, the NYU notes were called "Selected Topics in Geometry," reflecting the fact that the chief focus was on polyhedral rather than differentiable objects. Differential geometry, *per se*, is the concern of the middle two chapters. Both the isoperimetric inequality and the four-vertex theorem are given more than one proof. Also proved (although unnamed and without attribution, oddly enough) are the *Umlaufsatz* and Fenchel's Theorem: The total curvature of a closed space curve C is at least 2π, with equality if and only if C is a plane convex curve. Another result without name or attribution is the really basic Euler-Poincaré-Hopf Theorem: The Euler characteristic of a compact manifold is equal to the sum of the indices of any continuous vector field with isolated singularities.

It is quite remarkable how little overlap there is between the two parts of this volume. It is probably significant that such overlap as does occur is principally in the theory of vector fields and their relation to topology (PL in the first case, smooth in the second). Thus, in Part II the Poincaré (2-dimensional) part of the Euler-Poincaré-Hopf Theorem is given a pictorial proof, and then used to prove the Gauss-Bonnet Theorem: The total curvature of a compact surface equals its Euler characteristic times 2π. This theorem is without question the most famous result in all of global differential geometry. Its appeal seems rooted in the juxtaposition of a local geometric quantity, the Gauss curvature, with a global topological quantity, the Euler characteristic. (Incidentally, I have been asked, and have been unable to determine, who is responsible for attaching the label "Gauss-Bonnet" to this theorem; clearly neither Gauss nor Bonnet had any inkling of such a result.) Attempts to generalize this theorem have led to many of the basic advances in geometry in this century, including results of Hopf, Cohn-Vossen, Allendoerfer, Weil, Weyl, and Chern. I should add that an important feature of the Gauss-Bonnet Theorem is that on the one hand it is an intrinsic result, true for any compact two-dimensional Riemannian manifold, but on the other hand it gains added significance when applied to a surface immersed in $\mathbb{R}^3$, where the total curvature is determined by the degree of the Gauss map. (In particular, it is in that context that it is tied to the important work of Herman Weyl on the volume of tube domains.)

The remainder of Part II is devoted to global *extrinsic* results, those concerning compact or complete surfaces in $\mathbb{R}^3$, associated with the names of Hadamard, Hilbert, Liebmann, Cohn-Vossen, Hopf, and Alexandrov. This genre of result concerns the way the surface lies in space, for example as the boundary of a convex set, or as a standard sphere. In some important cases it asserts that a given abstract surface cannot lie in $\mathbb{R}^3$. The most famous of these is Hilbert's Theorem that there is no surface in $\mathbb{R}^3$ that serves as a model for the hyperbolic plane. One often sees the pseudosphere described as such a model, but although it has the same local properties, characterized by constant negative curvature, it represents only a small part of the whole hyperbolic plane (small, for example, in that it has finite area, whereas the whole hyperbolic plane has infinite area). Even the universal covering surface of the pseudosphere has at every point only one direction in which a geodesic ray can be extended arbitrarily far. In every other direction one reaches at finite distance a boundary point at which the surface becomes singular. One way to describe the nature of the singularity is to note that the mean curvature of the surface becomes infinite. It follows that there is no possible extension of the pseudosphere to a larger smooth surface. The key notion here is that of completeness, introduced by Hopf and Rinow. One of the various characterizations of completeness is that every geodesic may be extended arbitrarily far in both directions. That is certainly true of the hyperbolic plane. The version of Hilbert's Theorem proved here is that there does not exist a complete surface of constant negative curvature in $\mathbb{R}^3$.

Hilbert also gave a proof of a theorem originally due to Liebmann concerning surfaces of constant *positive* curvature. Hopf gives an updated form of Hilbert's proof, and shows that *there are no complete surfaces of constant positive curvature in* $\mathbb{R}^3$ *except for the standard sphere*. This particular result, aside from its intrinsic interest, is remarkable for the number of ways that it may be viewed.

First, one could write an entire book on the subject of characterizations of the sphere. The list would obviously include extremal properties, both mathematical (such as isoperimetric, or via eigenvalues) and physical (such as gravitational or electrostatic equilibrium figure), differential

geometric (such as Liebmann's Theorem above and others to be discussed below), and via symmetries. The list is also added to each year. (To cite just one example, see the article "Potato Kugel" by Aharonov, Schiffer, and Zalcman in Israel Journal of Mathematics, vol. 40 (1981), pp. 331–339.)

The second is as part of the theory of *ovaloids*: compact surfaces in $\mathbb{R}^3$ with positive curvature, bounding a convex body. Hilbert's proof actually assumes that the surface is an ovaloid, but a theorem of Hadamard (for which Hopf gives three different proofs in the notes) states that a compact surface in $\mathbb{R}^3$ with positive curvature is in fact an ovaloid. The theory of ovaloids is quite extensive, since the convexity assumption allows one to make many arguments not otherwise available.

The third aspect concerns *rigidity*. Note that in all of the above discussion the word "curvature" has referred to the *Gauss curvature* K of the surface. Gauss' famous *Theorem Egregium* states that K is an intrinsic quantity, and in particular is invariant under "bendings" of the surface, where lengths on the surface are unchanged and only its position in space is varied. Thus Liebmann's Theorem implies that a sphere is *rigid*; it cannot be "bent" into another shape, keeping lengths unchanged. Cohn-Vossen later showed that the rigidity of the sphere was a special case of the more general fact that all ovaloids are rigid. This result is clearly the smooth analog of Cauchy's Theorem from Part I stating that all convex polyhedra are rigid.

Next instead of considering the Gauss curvature, we could use the mean curvature H. Surfaces with H constant arise naturally as stationary values for the variational problem: minimize area, fixing the enclosed volume. Locally they arise physically as equilibrium surfaces separating two regions of different air pressures, as in soap bubbles. The theorem that compact surfaces of constant mean curvature are spheres was proved first by Liebmann for ovaloids, then by Hopf for arbitrary immersed surfaces of genus zero (that is, topologically a sphere), and then by Alexandrov for surface of arbitrary genus provided that they are *embedded* (that is, with no self-intersections). The proofs (all given in the notes) use three entirely different arguments.

Finally, both conditions: K constant and H constant are special cases of the more general condition that K and H are functionally related: $U(K, H) = 0$ for a suitable (smooth) function U. Surfaces satisfying such a condition are called *Weingarten surfaces*. It turns out that a number of theorems proved for surfaces of constant Gauss or mean curvature hold more generally for Weingarten surfaces, possibly with certain restrictions on the relation between H and K. The characterization of the sphere in terms of H or K being constant is a case in point.

These then are typical results in surface theory in the large, and their proofs make up the bulk of the material in Hopf's Stanford notes. In view of the fact that almost thirty years have elapsed between the original lectures and their publication (and almost forty in the case of the NYU notes) one must obviously address certain questions. Are the notes still relevant or are they hopelessly out of date? Where has the subject gone in the interim? What if any special qualities do the notes possess that might justify their publication at this late date?

To start with the subject, differential geometry in the large has grown from what may fairly be described as a narrow specialty at the time of the original lectures to its current place as a major subject of study, with important ties to many fields, including mathematical physics, dynamical systems and ergodic theory, analysis, and topology. Some aspects of the field are far removed from the matter treated in these lectures, but others are directly linked. Certainly one of the major effects of the Stanford notes was to diffuse more widely the methods of Hopf and of Alexandrov used to attack the basic question of whether the sphere is the only compact surface of constant mean curvature in $\mathbb{R}^3$. Both of those methods have been applied to a variety of other situations. The original question eluded all assaults (and there have been many) until just now, when it has apparently been settled by Henry Wente in a paper to appear in the Pacific Journal. Wente proves that there exists an immersed *torus* of constant mean curvature. (A Gallup poll of workers in the field would probably have resulted earlier in an overwhelming consensus that no such surface existed.)

The particular quality that makes this volume the classic that I believe it is, is that, as mentioned earlier, many results are given two and even three proofs. That seems to me emblematic of an approach whose goal is to impart insight and depth of understanding. It is the antithesis of the sparse, efficient march to the rhythm of "Theorem, Proof, Theorem, Proof." As I said at the outset, Hopf's combined depth of insight and clarity of presentation made his lectures an absolute delight. Fortunately, much of their flavor is preserved here for those in the present and future generations.

From Affine to Euclidean Geometry. An Axiomatic Approach. By Wanda Szmielew. Edited, prepared for publication and translated from the Polish by Maria Moszyńska, Warsaw University; PWN-Polish Scientific Publishers, Warszawa, Poland; D. Reidel Publishing Company, Dordrecht, Holland; Boston, MA; London, England, 1983, xii + 194 pp.

VICTOR PAMBUCCIAN
Str. Viitorului 26, 70266 Bucuresti, Romania

Descartes (1637) was the first to realize the real significance of algebra for geometry. In essence he observed that a "Cartesian plane" is a model of Euclidean geometry. It was more than 250 years later that Hilbert (1899) was able to prove the converse: *Any model of complete Euclidean geometry is isomorphic to the Cartesian plane constructed over the field of real numbers*. This was a first and significant step toward fully realizing the connections between geometry and algebra.

But Hilbert's axiom system is too much *Euclidean* to be simple. When formalized it needs three kinds of individual variables, for points, lines and planes, and four primitive notions (predicates), "incidence," "betweenness," "congruence of segments," "angle-congruence," and thus looks rather baroque.

The most important step toward simplifying the axiom system of Hilbert was taken by Tarski (who also supervised W. Szmielew's Ph.D. thesis). His axiom system is expressed in a "language" containing only one kind of individual variable (for "points") and two primitive notions "B" and "D", where "$B(abc)$" reads "point b lies between a and c" and "$D(abcd)$" reads "point b is as distant from a as point d is from c" (or, equivalently "segment ab is congruent to segment cd"). All its axioms are expressed in first-order logic and its representation theorem can be stated as follows:

Any model of Tarski's axiom system is isomorphic to a Cartesian plane constructed over a "real-closed field"

Still, algebra looks rather different from geometry. Algebraic axiom systems admit subsystems for weaker algebraic structures. For example, the axiom system for ordered fields admits subsystems for fields, quasi-fields, rings, and groups (commutative or not). As to geometry, Tarski does not provide an axiom system for any "geometry" whose coordinatization would lead to an algebraic structure more general than ordered (Pythagorean) fields.

To what extent can one construct corresponding algebraic and geometric structures simultaneously? What could be said about Euclidean geometry by a being which has not yet learned what measuring and comparing distances means, but is able to see that three points are collinear? Both questions were partially answered before. D. Barbilian (1940) constructed projective (and implicitly affine) geometry over rings, but had to restrict the class of rings in order to avoid geometric teratology. The answer to the second question is embodied by affine geometry. Now, given affine geometry one might further ask, what additional axioms are needed in order to introduce the order relation on a line and a "metric"? R. Nevanlinna (1976) gave a partial answer to this question.

The book of Wanda Szmielew answers for the first time all the questions thus raised. It presents the relationship between algebra and geometry at the earliest stage in the development of concepts, beginning with the weakest algebraic structures (which may be neither commutative nor associative) for which there are *reasonable* geometric equivalents. The key word for this relationship is "representation theorem". Indeed, there are a dozen representation theorems in this work, generalizing those of Hilbert and Tarski mentioned above.

In Szmielew's treatment, affine geometry is based on a quaternary relation among points called parallelity ("$ab\|cd$"), instead of being built up, as usual, on collinearity (a ternary relation "L", with "$L(abc)$" meaning "the points a, b, c are collinear"). These relations are mutually definable by:

$$ab\|cd \leftrightarrow \big(a = b \vee \forall p(L(abp) \rightarrow L(cdp)) \vee \forall p(L(abp) \rightarrow \neg L(cdp))\big)$$

and

$$L(abc) \leftrightarrow ab\|bc.$$

So, what is gained by this non-traditional approach to affine geometry? The resulting axiom system, for what the author calls "parallelity planes", is simpler, in the sense that parallelity planes admit an $\forall\exists$-axiom system, whereas affine geometry based on collinearity doesn't. (A $\forall\exists$-axiom system is one in which all the axioms are written in prenex form so that all universal quantifiers precede all existential quantifiers.)

The coordinatization of parallelity planes leads to a very weak algebraic structure called a "ternary field". If one adds the minor or major Desargues theorem then one gets quasi-fields or skew fields, respectively, as coordinate algebraic structures. If one adds instead Pappus' law the coordinate field becomes commutative.

In this way algebra and geometry develop in parallel. The odyssey of travelling to Euclidean geometry by way of affine geometry is the content of this marvellous book. Unfortunately, Ithaca is not reached. The author died before finishing the last chapter. Most gaps have been filled by the editor, M. Moszyńska, whose style is very close to that of the author.

This book leaves the reviewer with a feeling such as one experiences watching the throwing of the javelin at the Olympic Games. Every movement seems simple, natural, the way it has to be. One almost forgets that behind this apparent ease and grace there are years of hard work, of sweat and tears.

The present book is a life's work. As expected from an accomplished scholar, the language is simple and precise. The monograph is suitable for a beginning graduate course in the foundations of geometry.

The bibliography of the works of Wanda Szmielew at the end of the book is there to tell us that she praised above all in mathematics meaningfulness and beauty.

ANSWERS TO REBUSES ON PAGES 56–58

1. Counterexample.

2. Cross product.

3. $\Gamma(\frac{1}{2})$. (See p. 76.)

4. A group of translations of the plane.

5. Orthogonal matrix.

6. Scalar matrix.

7. Heaviside function.

LETTERS TO THE EDITOR

Material for this department should be prepared exactly the same way as submitted manuscripts (see the inside front cover) and sent to Professor P. R. Halmos, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053.

Editor:

Harary and Robinson's note [**1**] was quite timely for me, since I had just set my Discrete Mathematics class a takehome examination problem to investigate how the diameter of a graph can be related to the diameter of its complement. It might be of interest to note that the complete answer to this problem follows from some observations of my students, together with Harary and Robinson's theorem and a strengthening of that theorem.

Recall some definitions. If G is a graph, its *complement* $\overline{G}$ has the same vertex set, and has edges precisely where G does not have edges. The *distance* $d_G(u,v)$ between two vertices u and v of a graph G is the length of the shortest path from u to v, with the convention that $d_G(u,v)=\infty$ if there is no such path. The *diameter* $d(G)$ of G is the maximal distance between two of its vertices.

Harary and Robinson proved that if $d(G) \geqslant 3$, then $d(\overline{G}) \leqslant 3$. A modification of their reasoning establishes the following result.

THEOREM. *If* $d(G) \geqslant 4$, *then* $d(\overline{G}) \leqslant 2$.

Proof. Let u and v be vertices of G such that $d_G(u,v) \geqslant 4$, and let x be any other vertex of G. We observe that, in G, if x is adjacent to v, then x cannot be adjacent to u, or to any vertex adjacent to u (or there would be a path from u to v of length less than 4). Translating to the complement, we get

(*) in $\overline{G}$, if x is not adjacent to v, then x must be adjacent to u, and to any vertex not adjacent to u.

Now let x and y be any two vertices in $\overline{G}$. They are both adjacent to either u or v, and if they are either both adjacent to u or both adjacent to v, then $d_{\overline{G}}(x,y) \leqslant 2$. On the other hand, if, say, x is adjacent to u but not to v, and y is adjacent to v but not to u, then (*) implies that x must be adjacent to y. Hence $d_{\overline{G}}(x,y) \leqslant 2$ and $d(\overline{G}) \leqslant 2$. ■

These results imply that the following is a complete list of unordered pairs possible for the diameter of a graph and its complement.

Pair	$\{0,0\}$	$\{1,\infty\}$	$\{2,n\}$ $(n \geqslant 2)$	$\{2,\infty\}$	$\{3,3\}$
Example	P_1	P_2	C_{2n+1}	P_3	P_4

Here P_k is the path graph on k vertices, C_k is the cycle graph on k vertices.

Reference

1. F. Harary and R. W. Robinson, The diameter of a graph and its complement, this MONTHLY, 92 (1985) 211–212.

Philip D. Straffin, Jr.
Mathematics and Computer Science
Beloit College
Beloit, WI 53511

REBUS # 3, PAGE 75

The answer is really "square root of π", or $\sqrt{\pi}$, which, as everybody knows, is the value of the Gamma function at the point $\frac{1}{2}$.

The Mathematical Association of America

CORPORATE MEMBERS

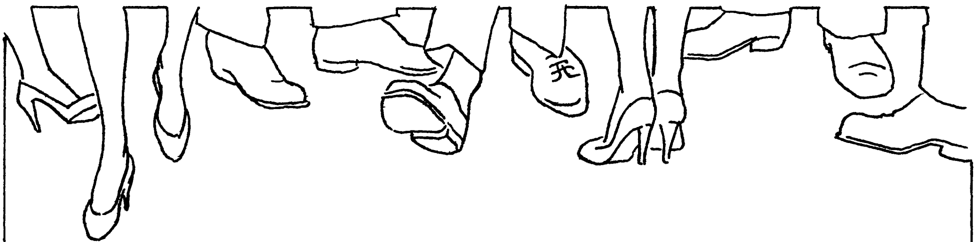
The Mathematical Association of America gratefully acknowledges the interest and support of the following:

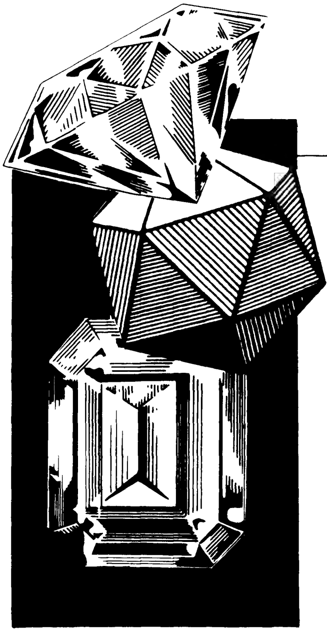
CORPORATE MEMBERS

Addison-Wesley Publishing Company, Inc.
American Mathematical Society, Inc.
Birkhäuser Boston, Inc.
Digital Equipment Corporation
International Business Machines Corporation
Macmillan Publishing Company
Marcel Dekker, Inc.
The Mitre Corporation
Pergamon Press, Inc.
Prindle Weber and Schmidt, Inc.
Saunders College Publishing, Inc.
Scott Foresman & Company, Inc.
Springer-Verlag New York, Inc.
John Wiley & Sons, Inc.

CONTRIBUTING CORPORATE MEMBER

Bell Laboratories, Inc.





Mathematical Gems III

by Ross Honsberger

260 pp. Hardbound, 1985.

ISBN-0-88385-313-2

MAA Member: $21.00

List: $27.00

This book is packed with gems from elementary combinatorics, geometry, number theory, and probability. It will bring the enjoyment of good mathematics to the student as well as the teacher.

Glance at the Table of Contents and order now!

Order from:

The Mathematical Association of America

1529 Eighteenth Street, N.W.

Washington, D.C. 20036

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 93, Number 2 February 1986

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. (This is necessary for indexing purposes.) Send all main articles, letters to the editor, and editorial correspondence to the editor. Contributions to the other four special departments (Notes, Teaching of Mathematics, Unsolved Problems, and Problems and Solutions) should be sent to the appropriate one of the associate editors, whose addresses are given below.

Notes. Combinatorics and probability: SABRA S. ANDERSON, Department of Mathematics, University of Minnesota, Duluth, MN 55812. Analysis and linear algebra: SHELDON AXLER, Department of Mathematics, Michigan State University, East Lansing, MI 48824. Topology, geometry, and abstract algebra: J. ARTHUR SEEBACH, JR., Department of Mathematics, St. Olaf College, Northfield, MN 55057. Authors of fields not covered by these rubrics may send their notes to any one of the three editors.

Teaching of Mathematics: ROBERT F. WARDROP, Department of Mathematics, Central Michigan University, Mt. Pleasant, MI 48859.

Unsolved Problems: RICHARD GUY, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.

Proposed problems and solutions, both Elementary and Advanced: D. H. MUGLER, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053.

Reprint Permission: A. B. WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Back Issues: P. and H. BLISS Co., Middletown, CT 06457.

Microfilm Editions: University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

Editorial Assistants: Fanny Hailpern, Mary P. Jackson.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT, ten times a year: January, February, March, April, May, June-July, August-September, October, November and December.

The annual subscription price for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is \$24 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are \$24. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is \$70 per year.

Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

PRINTED IN THE UNITED STATES OF AMERICA

SOME CONSEQUENCES OF THE STURM COMPARISON THEOREM

ANDREA LAFORGIA
Department of Mathematics, University of Torino, Via Carlo Alberto 10, 10123 Torino, Italy

MARTIN E. MULDOON
Department of Mathematics, York University, Downsview, Ontario M3J 1P3, Canada

To Lee Lorch on his seventieth birthday

1. Introduction. It is almost 150 years since the appearance of the Sturm comparison theorem [**22**] concerning solutions of second order linear differential equations. The theorem has been greatly extended and generalized, most notably to higher order equations, to systems and to partial differential equations; see [**20**], [**23**]. Elementary textbooks, e.g. [**5**], [**12**], [**21**], discuss what is essentially Sturm's original result and some of its more immediate applications. The purpose of this expository paper is to point to some further applications which deserve to be better known and which can be discussed in beginning courses. We also hope to convince the reader of the power of the Sturm comparison theorem, even in its original form, in providing information about zeros of special functions.

Among the special functions to be discussed here is the so-called Bessel function of the first kind [**27**]

$$J_\nu(x) = \left(\frac{x}{2}\right)^\nu \sum_{k=0}^{\infty} \frac{(-1)^k (x/2)^{2k}}{k!\Gamma(\nu+k+1)}$$

which satisfies

$$x^2 y'' + xy' + (x^2 - \nu^2) y = 0. \tag{1}$$

$J_\nu(x)$ has an infinite sequence of positive x-zeros $j_{\nu k}$, $k = 1, 2, \ldots$, which are intimately connected with the eigenvalues of some important boundary value problems such as that arising from the vibration of a circular membrane [**6**, Chap. 5]. We will have occasion also to mention the ultraspherical polynomials [**25**, Chap. 4]

$$P_n^{(\lambda)}(x) = \sum_{m=0}^{[n/2]} (-1)^m \frac{\Gamma(n-m+1)}{\Gamma(\lambda)\Gamma(m+1)\Gamma(n-2m+1)} (2x)^{n-2m},$$

$\lambda > -\frac{1}{2}$, which are orthogonal on $[-1,1]$ with respect to the weight function $(1-x^2)^{\lambda - 1/2}$. $P_n^{(\lambda)}(x)$ is the polynomial solution of the differential equation [**25**, p. 80]

$$(1 - x^2) y'' - (2\lambda + 1) xy' + n(n + 2\lambda) y = 0 \tag{2}$$

which is normalized by the condition

$$P_n^{(\lambda)}(1) = \binom{n + 2\lambda - 1}{n}.$$

In the special case $\lambda = \frac{1}{2}$, we have the Legendre polynomials, and in the cases $\lambda = 0, 1$, we have the Chebyshev polynomials of the first and second kind, respectively. These are among the special

---

Andrea Laforgia studied mathematics at the University of Turin, Italy. Having spent a year at the Centro di Calcolo, University of Genoa, he returned in 1974 to Turin where he is now Associate Professor of Mathematics. His mathematical interests are in special functions and numerical analysis; he plays bridge in his spare time.

Martin Muldoon studied at University College Galway, Ireland, and at the University of Alberta where he received a Ph.D. in Mathematics under the supervision of Lee Lorch. He is now Professor of Mathematics at York University. He is interested in various aspects of classical analysis including special functions and differential equations.

functions whose zeros are important in numerical integration [**1**, Chap. 25; **25**, Chap. 15]. The zeros of the ultraspherical polynomials $P_n^{(\lambda)}(x)$, all in $(-1,1)$, can be interpreted as the one-dimensional positions of equilibrium of n ($\geqslant 2$) unit charges in the interval $(-1,1)$ in the field generated by two identical charges of magnitude $\frac{\lambda}{2}+\frac{1}{4}$ placed at 1 and -1. (See [**25**, pp. 140–142].)

Finally we remark that besides Sturm methods and differential equations there are many other approaches to the study of zeros of special functions [**19**], [**25**], [**27**].

2. The theorem. We now present the Sturm comparison theorem in as simple a form as possible but one which will be adequate for most of our purposes. We suppose that y and Y satisfy

$$y''+f(x)y=0,$$
$$Y''+F(x)Y=0,$$

respectively, where f and F are continuous on an interval $[a,b]$. Multiplying the first equation by Y, the second by y, subtracting and integrating we get

$$[y'Y-yY']_x^z=\int_x^z(F-f)\,yY\,dx. \tag{3}$$

Thus

(i) *If $y(a)=Y(a)=0$, $y'(a+)=Y'(a+)>0$ and $F(x)>f(x)$, $a<x<b$, then $y(x)>Y(x)$, $a<x\leqslant c$, where c is the first zero of Y on $(a,b]$ to the right of a.*

This follows by taking $x=a$ in (3) and noting that y/Y increases from 1 as x increases. From (i) we get:

(ii) *If $y(a)=Y(a)=0$ and $F(x)>f(x)$, $a<x<b$, then the first zero of Y to the right of a on $(a,b]$ occurs before the first zero of y to the right of a.*

Finally we can claim:

(iii) *Under the hypotheses of (ii), the kth zero of Y to the right of a on $(a,b]$ occurs before the kth zero of y to the right of a.*

This is most easily seen by observing that between each pair of zeros of y there is a zero of Y. We can suppose, without loss of generality, that

$$y(x_1)=y(x_2)=0,\quad y'(x_1)>0,\quad y'(x_2)<0,\quad Y(x)>0,\quad x_1\leqslant x\leqslant x_2;$$

we get a contradiction by using (3) with $x=x_1$, $z=x_2$.

These results are often stated for general self-adjoint second order ordinary differential equations; see, e.g., [**9**]. Various modifications have been made allowing for the possibility of a singularity in the differential equations at $x=a$; see [**16**], [**24**], [**25**]. Such a modification will be needed in §5. Moreover, the condition $f(x)<F(x)$ is not necessary for the conclusions and various stronger forms of the theorem have been proved [**10**], [**13**], [**14**], [**15**].

3. Application to a single equation; convexity of zeros. One easy consequence of the Sturm theorem is an application to the shapes of the successive arches of the graph of a nontrivial solution of a single differential equation. This result, as far as its application to zeros is concerned, was already in Sturm's paper [**22**, p. 173]; it is called the "Sturm convexity theorem" in [**3**, p. 272]. It was discussed also by E. Hille [**8**] and by G. Szegő [**24**] but finds its most detailed exposition in a paper by E. Makai [**17**]. We suppose that y is a nontrivial solution of

$$y''+\varphi(x)y=0, \tag{4}$$

where φ is continuous and decreasing on (a,b) and y has consecutive zeros at x_1,x_2,x_3 on

(a, b). Then

$$x_2 - x_1 < x_3 - x_2, \tag{5}$$

i.e., the sequence of zeros of y is *convex*. Similarly, if φ increases, the sequence is concave. The inequality (5) follows from the more general result that if we rotate the arch of the graph of y between x_1 and x_2 through 180° about x_2, the resulting arch will be contained entirely within the arch joining x_2 to x_3. (See Fig. 1.) To prove this last result we need to show that

$$y(x) > -y(2x_2 - x), \quad x_2 < x < x_2 + d,$$

where $d = x_2 - x_1$, and we suppose, without loss of generality, that $y(x) < 0$, $x_1 < x < x_2$.

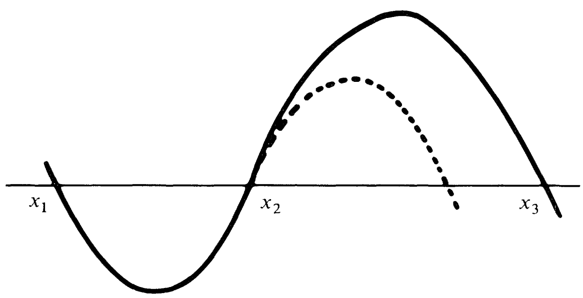


FIG. 1

Then we use the fact that $Y(x) = -y(2x_2 - x)$ satisfies

$$Y'' + \varphi(2x_2 - x)Y = 0,$$

so that the desired result follows from (i). Makai [**17**] points out that this method also gives the increase of the maxima of $|y(x)|$ under the stated hypotheses, a result known as "Sonin's theorem".

4. Spacing of zeros; variation with a parameter. An obvious consequence of the Sturm theorem, already noted in [**22**, pp. 167–169], achieved by comparing $y'' + f(x)y = 0$ with $y'' + m^2y = 0$ and $y'' + M^2y = 0$, is that if f is continuous on (a, b) and satisfies $m^2 < f(x) < M^2$, where m and M are positive numbers, then the zeros $x_1, x_2, \ldots,$ in increasing order, of y on (a, b) satisfy

$$\pi/M < x_{i+1} - x_i < \pi/m, \quad i = 1, 2, \ldots .$$

Our object in this section is to see how this spacing of the zeros changes with some parameter in, e.g., the case of orthogonal polynomials. This is related to the Sturm theorem by considering two equations for close values of the parameter and shifting one of them so that a zero of one solution coincides with a zero of the other. To our knowledge, the first use of this idea is contained in the work of G. Szegő and P. Turán [**26**] on zeros of ultraspherical polynomials. The main idea can be conveyed by considering a special case, the θ-zeros of Legendre polynomials. It follows from (2) that

$$y_n(\theta) = (\sin\theta)^{1/2} P_n(\cos\theta)$$

satisfies [**25**, p. 81]

$$y'' + \varphi_n(\theta)y_n = 0,$$

where

$$\varphi_n(\theta) = \left(n + \frac{1}{2}\right)^2 + (4\sin^2\theta)^{-1},$$

and now primes denote derivatives with respect to θ. For $n \geqslant 2$, the zeros of $P_n(\cos\theta)$ or of $y_n(\theta)$

on $(0,\pi)$ are denoted in increasing order by $\theta_1^{(n)},\dots,\theta_n^{(n)}$. These zeros are distributed symmetrically about $\theta=\pi/2$. An easy application of the Sturm theorem gives

$$\theta_k^{(n)} > \theta_k^{(n+1)}, \quad k=1,\dots,n.$$

We will show, as in [**26**], that

$$\theta_{k+1}^{(n+1)} - \theta_k^{(n+1)} < \theta_{k+1}^{(n)} - \theta_k^{(n)}, \quad k=1,2,\dots,\left[\frac{n-1}{2}\right], \tag{6}$$

in other words, the spacing of the zeros on $(0,\pi/2)$ decreases as n increases. Fig. 2 shows a sketch of the graphs of $y_4(\theta)$ and $y_5(\theta)$ for $0<\theta<\pi/2$; (6) says, in particular, that $AC < BD$.

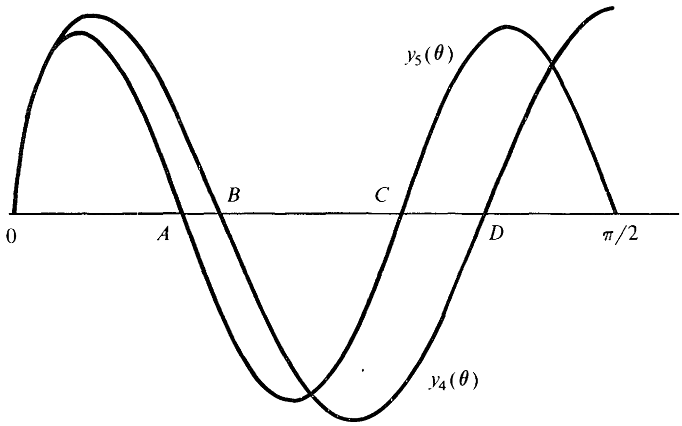


FIG. 2

To prove (6), we choose, for fixed k,

$$\delta = \theta_k^{(n)} - \theta_k^{(n+1)} > 0.$$

We use the fact that $y_n(\theta+\delta)$ satisfies

$$y'' + \varphi_n(\theta+\delta)y = 0$$

and it has consecutive zeros at $\theta_1^{(n)}-\delta,\ \theta_2^{(n)}-\delta,\dots$. This makes the kth zero of $y_{n+1}(\theta)$ coincide with the kth zero of $y_n(\theta+\delta)$. Since, obviously,

$$\varphi_n(\theta+\delta) < \varphi_n(\theta) < \varphi_{n+1}(\theta), \quad 0<\theta\leqslant \pi/2-\delta,$$

we see, from (ii), that the next larger zero of $y_{n+1}(\theta)$ occurs before the next larger zero of $y_n(\theta+\delta)$, that is

$$\theta_{k+1}^{(n)} - \delta > \theta_{k+1}^{(n+1)},$$

and so (6) holds.

The idea of this section has been applied to Laguerre polynomials in [**16**] and to some Jacobi polynomials in [**2**]. In the latter paper it is refined to give results on the increase as a function of n of the spacings of the zeros multiplied by a linear function of n.

5. An indirect use of the Sturm theorem; zeros of Bessel functions. There are many proofs (see [**19**] for references) that, with the notation of §1, each $j_{\nu k}$ increases as ν increases, $\nu\geqslant 0$. One of particular interest here is that due to M. B. Porter and M. Bôcher [**4**], which uses the Sturm theorem. On the other hand, $j_{\nu k}/\nu$ decreases as ν increases, $0<\nu<\infty$. Makai [**18**] used the Sturm theorem in an ingenious indirect way to show this. For this application we have to note that

the continuity of f and F *at* $x = a$ in §1 may be dispensed with provided we suppose that

$$\lim_{x \to a+} [y(x)Y'(x) - y'(x)Y(x)] = 0. \tag{7}$$

Makai uses the fact that the function

$$y_\nu(x) = x^{1/2} J_\nu\left(j_{\nu k} x^{1/(2\nu)}\right)$$

satisfies

$$y_\nu'' + p_\nu(x) y_\nu = 0,$$

where

$$p_\nu(x) = \left\{ j_{\nu k}/(2\nu) x^{1/(2\nu) - 1} \right\}^2.$$

This follows from (1) by suitable changes of variable.

Now suppose that for some μ and ν, with $0 < \mu < \nu$, we have

$$j_{\mu k}/\mu \leqslant j_{\nu k}/\nu. \tag{8}$$

The functions $y_\mu(x)$ and $y_\nu(x)$ both vanish at $x = 0$ and both have their kth positive zeros at $x = 1$. Moreover, they satisfy the condition corresponding to (7). Also from (8), $p_\mu(x) < p_\nu(x)$, $0 < x < 1$, so an application of (iii) shows that the kth zero of $y_\nu(x)$ occurs *before* the kth zero of $y_\mu(x)$. This contradiction shows that (8) cannot hold and so $j_{\nu k}/\nu$ decreases as ν increases, $0 < \nu < \infty$.

Makai's method has been used by Á. Elbert [7] to get even stronger results on zeros of Bessel functions.

It may be remarked that the most obvious direct approach of scaling the independent variable, i.e., considering the equation

$$y'' + \left[\nu^2 + \left(\frac{1}{4} - \nu^2\right) x^{-2}\right] y = 0,$$

(again got from (1)), satisfied by $x^{1/2} J_\nu(\nu x)$ is ineffective here since the Sturm theorem cannot be applied. Nevertheless the idea of scaling is sometimes useful. For example, in [**11**], it is used to show that $\lambda x_{nk}^{(\lambda)}$ increases with λ, $0 < \lambda < 1$, $k = 1, 2, \ldots, [n/2]$, where $x_{nk}^{(\lambda)}$ denotes the kth positive zero of the ultraspherical polynomial $P_n^{(\lambda)}(x)$.

The authors' research is supported by grants from Consiglio Nazionale delle Ricerche (Italy) and the Natural Sciences and Engineering Research Council (Canada).

References

1. M. Abramowitz and I. A. Stegun, editors, Handbook of Mathematical Functions, National Bureau of Standards Applied Mathematics Series 55, Washington, DC, 1964.

2. S. Ahmed, A. Laforgia, and M. E. Muldoon, On the spacing of the zeros of some classical orthogonal polynomials, J. London Math. Soc., (2) 25 (1982) 246–252.

3. G. Birkhoff and G.-C. Rota, Ordinary Differential Equations, 3rd ed., Wiley, New York, 1978.

4. M. Bôcher, On certain methods of Sturm and their application to the roots of Bessel's functions, Bull. Amer. Math. Soc., 3 (1897) 205–213.

5. J. C. Burkill, The Theory of Ordinary Differential Equations, 2nd ed., Oliver and Boyd, Edinburgh and London, 1962.

6. R. Courant and D. Hilbert, The Methods of Mathematical Physics, vol. 1, 1st English edition, Wiley-Interscience, New York, 1953.

7. Á. Elbert, Concavity of the zeros of Bessel functions, Studia Sci. Math. Hungar., 12 (1977) 81–88.

8. E. Hille, Über die Nullstellen der Hermiteschen Polynome, Jahresber. Deutsch Math.-Verein., 44 (1933) 162–165.

9. E. L. Ince, Ordinary Differential Equations, Longmans, London, 1926; reprinted Dover, New York, 1956.

10. K. Kreith, Comparison theorems for non-self-adjoint differential equations based on integral inequalities, Proc. Amer. Math. Soc., 34 (1972) 105–109.

11. A. Laforgia, A monotonic property for the zeros of ultraspherical polynomials, Proc. Amer. Math. Soc., 83 (1981) 757–758.

12. W. Leighton, A First Course in Ordinary Differential Equations, 5th ed., Wadsworth, Belmont, CA, 1981.

13. ______, A substitute for the Picone formula, Bull. Amer. Math. Soc., 55 (1949) 325–328.

14. ______, Comparison theorems for linear differential equations of the second order, Proc. Amer. Math. Soc., 13 (1962) 603–610.

15. A. Yu. Levin, A comparison principle for second-order differential equations, Soviet Math. Dokl. 1 (1960) 1313–1316. [Russian original, Dokl. Akad. Nauk. SSSR, 136 (1960) 783–786.]

16. L. Lorch, Elementary comparison techniques for certain classes of Sturm-Liouville equations. (Proc. Internat. Conf., Uppsala, 1977), Sympos. Univ. Upsaliensis Ann. Quingentesimum Celebrantis 7 (Almqvist and Wiksell, Stockholm, 1977) pp. 125–133.

17. E. Makai, On a monotonic property of certain Sturm-Liouville functions, Acta. Math. Acad. Sci. Hungar., 3 (1952) 165–172.

18. ______, On zeros of Bessel functions, Univ. Beograd Publ. Elektrotehn. Fak. Ser. Mat. Fiz., no. 602–633 (1978) 109–110.

19. M. E. Muldoon, The variation with respect to order of zeros of Bessel functions, Rend. Sem. Mat. Univ. Politec. Torino, 39 (1981) 15–25.

20. W. T. Reid, Sturmian Theory for Ordinary Differential Equations, Applied Mathematical Sciences, vol. 31, Springer-Verlag, New York, 1980.

21. G. F. Simmons, Differential Equations with Applications and Historical Notes, McGraw-Hill, New York, 1972.

22. C. Sturm, Mémoire sur les équations différentielles du second ordre, J. Math. Pures Appl., 1 (1836) 106–186.

23. C. A. Swanson, Comparison and oscillation theory of linear differential equations, Academic Press, New York and London, 1968.

24. G. Szegő, Inequalities for the zeros of Legendre polynomials and related functions, Trans. Amer. Math. Soc., 39 (1936) 1–17; Collected Papers, vol. 2 (Birkhaüser, 1982) 593–609.

25. ______, Orthogonal Polynomials, Amer. Math. Soc. Colloquium Publications, vol. 23, 4th ed., 1975.

26. G. Szegő and P. Turán, On the monotone convergence of certain Riemann sums, Publ. Math. Debrecen, 8 (1961) 326–335.

27. G. N. Watson, A Treatise on the Theory of Bessel functions, 2nd ed., Cambridge University Press, 1944.

A GEOMETRIC DERIVATION OF THE DISTRIBUTION OF THE CORRELATION COEFFICIENT $|r|$ WHEN $\rho = 0$

WILLIAM A. CHANCE

School of Business and Public Administration, University of Missouri-Kansas City, 5100 Rockhill Road, Kansas City, MO 64110

Introduction. R. A. Fisher [2] was among the first statisticians to discuss the geometric properties of bivariate data sets. Treating each data set of x and y as a point on an n-dimensional sphere, he noted that the correlation coefficient corresponds to the cosine of the angle between the radii to the two points. In the remainder of the paper, however, Fisher relied upon an analytical treatment of the bivariate normal distribution to derive a general expression for the frequency distribution of r.

Several later writers have appealed to geometric arguments, but in most cases they have reverted to an analytical approach to present their results. Jackson [**6**] and Hotelling [**4** and **5**] are examples. One reason for preferring an analytic approach may have been revealed by Hotelling when he indicated that "... a sketch is included of a modification of Fisher's derivation of the distribution of r designed to appeal to those who distrust geometrical arguments" [**5**, p. 194].

There have been exceptions to this reluctance to use geometric arguments to arrive at statistical

William A. Chance: I received my Ph.D. in Economics in 1964 at the University of Kansas, where I was persuaded to take a minor in mathematics. Since then, I have accorded mathematics and mathematicians the reverence they so justly deserve. I find recreation in tennis, fencing, swimming and grandfathering.

results. In supporting their efforts to show that familiar results in the theory of estimation can be represented as properties of linear spaces, Durbin and Kendall stated: "In the ultimate analysis geometrical 'proofs' in more than three dimensions are only restatements of analytical results in a special language; but they are nevertheless very useful, partly because of their elegance and partly because they carry a greater degree of conviction and understanding, to some minds at least, than the analytical approach" [**1**, p. 150].

This paper presents by geometric arguments a method of calculating directly probabilities for the sample correlation coefficient on the assumption that the sample is drawn from a bivariate population in which the correlation is $\rho = 0$. The vectors representing the sample data are assumed to have a spherically symmetric distribution. The required probabilities are shown to be the ratio of two volumes in a space containing the vectors of observed values of the two variables.

The geometry of correlation. Let $\mathbf{y}$ and $\mathbf{x}$ be column vectors, the elements of which are deviations from mean of the observations in y and x. Then the Pearson correlation coefficient can be expressed as

$$r = \mathbf{y}'\mathbf{x}/(\mathbf{y}'\mathbf{y})^{1/2}(\mathbf{x}'\mathbf{x})^{1/2},$$

where $\mathbf{y}'\mathbf{x}$, for example, is the standard inner product of the vectors. But this expression is also the value of the cosine of the angle between the vectors. Thus, if u is the angle between the vectors and $0 < u < \pi/2$, then $r = \cos u$.

If there are n_0 observations of y and x, the maximum dimensionality of the space required to contain the vectors will be $n_0 - 1$.

To show this result, let X be the matrix of order $n_0 \times 2$ whose columns are the vectors $\mathbf{y}$ and $\mathbf{x}$, $X = (\mathbf{y}, \mathbf{x})$. Consider the case $n_0 = 2$. Then X contains two row vectors; and it is evident that each row vector is the other multiplied by -1, because the sum of the elements in each column is zero. Therefore, there is at most only one linearly independent row vector among the two rows. It follows that the vectors will require at most a one-space, which is the trivial case in which $\mathbf{y}$ and $\mathbf{x}$ lie on the same line. Then $|\mathbf{r}| = 1$.

In the case of $n_0 = 3$, the sum of the row vectors in X must equal the zero vector. Thus, there are at most two linearly independent rows and the column vectors $\mathbf{y}$ and $\mathbf{x}$ require at most a

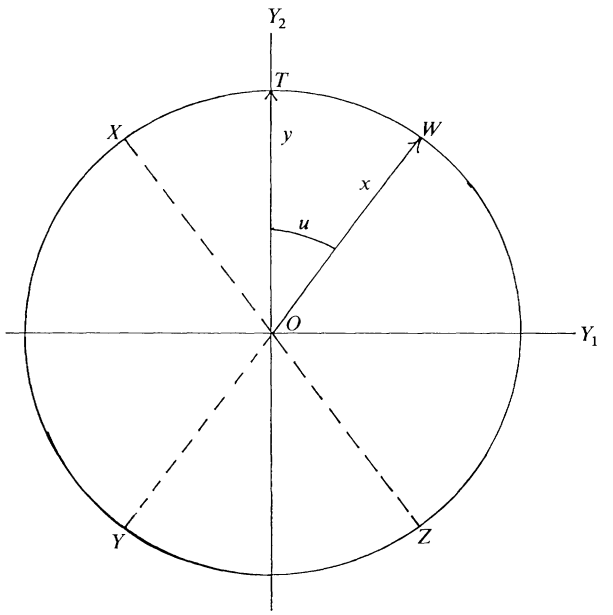


FIG. 1. Two vectors in a two-space.

two-space to contain them. Generally, the vectors **y** and **x** of order $n_0 \times 1$ will require at most an $(n_0 - 1)$-space to contain them.

The angle between two vectors does not depend on their lengths, and we can normalize the vectors to have unit length.

The case $n_0 = 2$ is trivial, and we consider the case $n_0 = 3$. Assume that the two vectors require a two-space. With no loss of generality, the vector **y** can be considered as aligned along an axis Y_2, as shown in Fig. 1. Then the circle in Fig. 1 traces the possible positions that could be taken by **x** relative to **y**. The angle between **y** and **x** is $u = \cos^{-1} r$.

Assume the vectors **x** and **y** are sample data randomly drawn from a bivariate population in which the variables x and y are uncorrelated: $\rho = 0$. Assume also that all directions taken by **x** and **y** in the two-space are equally probable, so that they have a spherically symmetric distribution. Let r_c denote the calculated value of r in the sample drawn. Then

$$P = P(|r| \geq r_c)$$

is the ratio of the area contained in the triangular sectors OWX and OYZ to the total area of the circle in Fig. 1. This is simply the ratio of area OWT to the area of the first quadrant of the circle.

Recall that **y** and **x** have been normalized to length 1. Then the area of the circle contained in the first quadrant is

$$A_1 = \int_0^{\pi/2} \int_0^1 a\, da\, du = \pi/4.^*$$

Now let u_c denote the angle between **y** and **x** corresponding to r_c. Then the area of OWT with angle u_c is

$$A_2 = \int_0^{u_c} \int_0^1 a\, da\, du = u_c/2.$$

The required probability is

$$P = A_2/A_1 = (2/\pi)\, u_c = (2/\pi)\cos^{-1} r_c. \tag{1}$$

For $n_0 = 4$ observations, the two vectors will require at most a three-space. Considering **y** fixed on an axis, the possible positions of **x** relative to **y** generate a ball with radius 1. Fig. 2 illustrates the situation. As before, the required probability will be $P = V_2/V_1$, where V_1 is the volume of the ball contained in the first octant and V_2 is the volume, contained in the first octant, of the conical section generated by rotating **x** around **y** with a fixed angular displacement u.

The volume of the ball in Fig. 2, contained in the first octant, is

$$V_1 = \int_0^{\pi/2} \int_0^{\pi/2} \int_0^1 a^2 \sin u\, da\, du\, dv = \pi/6.$$

The volume of the conical section of Fig. 2, contained in the first octant, is

$$V_2 = \int_0^{\pi/2} \int_0^{u_c} \int_0^1 a^2 \sin u\, da\, du\, dv = (\pi/6)(1 - \cos u_c).$$

Hence

$$P = V_2/V_1 = (\pi/6)(1 - \cos u_c)/(\pi/6) = 1 - \cos u_c,$$

and

$$P = 1 - r_c. \tag{2}$$

This development can now be generalized. Suppose that the two vectors **y** and **x** are of n_0

---

*Polar coordinates are used with a replacing the more commonly used r, in order to avoid confusion with the correlation coefficient.

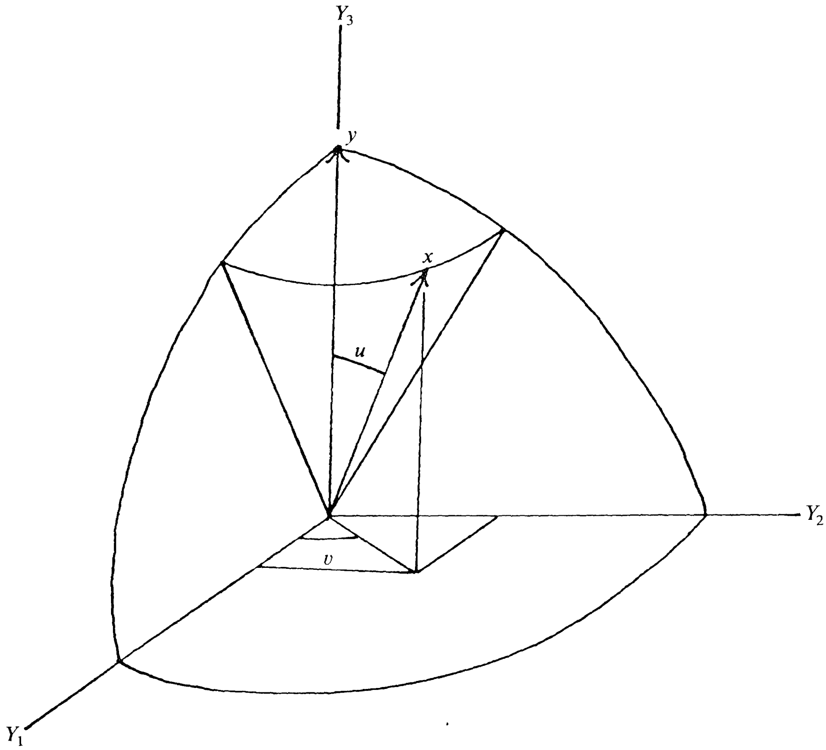


FIG. 2. Two vectors in a three-space.

observations contained in a vector space of dimension $n_0 - 1$. For $n = n_0 - 3$, the volumes required are

$$V_1 = \int_0^{\pi/2} \cdots \int_0^1 a^{n+1} \sin^n u \sin^{n-1} v_1 \cdots \sin v_{n-1}\, da\, du\, dv_1 \cdots dv_n$$

and

$$V_2 = \int_0^{\pi/2} \cdots \int_0^{u_c} \int_0^1 a^{n+1} \sin^n u \sin^{n-1} v_1 \cdots \sin v_{n-1}\, da\, du\, dv_1 \cdots dv_n .$$

Then

$$P = V_2/V_1 = \int_0^{u_c} \sin^n u\, du \Big/ \int_0^{\pi/2} \sin^n u\, du . \tag{3}$$

The form taken by (3) depends on whether the number of observations is even or odd. For n_0 even,

$$P = 1 - r_c - r_c(1 - r_c^2) \sum_{\substack{s=3 \\ s \text{ odd}}}^{n} \frac{1 \cdot 3 \cdot 5 \cdots (s-2)}{2 \cdot 4 \cdot 6 \cdots (s-1)} (1 - r_c^2)^{(s-3)/2}, \tag{4}$$

where n is odd, $n \geq 3$. For $n = 1$, we have $P = 1 - r_c$. For n_0 odd,

$$P = \frac{2}{\pi} \left\{ \cos^{-1} r_c - r_c(1 - r_c^2)^{1/2} \left\{ 1 + \sum_{\substack{s=2 \\ s \text{ even}}}^{n} \frac{2 \cdot 4 \cdots (s-2)}{3 \cdot 5 \cdots (s-1)} (1 - r_c^2)^{(s/2-1)} \right\} \right\}, \tag{5}$$

n is even, $n \geq 2$. For $n = 0$, $P = (2/\pi)\cos^{-1} r_c$.

Conclusion. The geometric method generates the correct distribution of $|\mathbf{r}|$ when $\rho = 0$. As a matter of fact, it is identical to the frequency distribution of Fisher [**3**, p. 657]:

$$df = \frac{\Gamma((n_0 - 1)/2)}{\Gamma((n_0 - 2)/2)\Gamma(1/2)}(1 - r^2)^{(n_0-4)/2}\,dr.$$

For this density function the value of P in equation (3) is

$$P = \int_r^1 (1 - r^2)^{(n_0-4)/2}\,dr \Big/ \int_0^1 (1 - r^2)^{(n_0-4)/2}\,dr.$$

Substitute $r = \cos u$, $n = n_0 - 3$ and equation (3) follows.

References

1. J. Durbin and M. G. Kendall, The geometry of estimation, Biometrika, 38 (1951) 150–158.

2. R. A. Fisher, Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population, Biometrika, 10 (1915) 507–521.

3. ______, The general sampling distribution of the multiple correlation coefficient, Proc. Roy. Soc. London, 121 (1928) 654–673.

4. Harold Hotelling, Tubes and spheres in n-space, and a class of statistical problems, Amer. J. Math., 61 (1939) 440–460.

5. ______, New light on the correlation coefficient and its transforms, J. Roy. Statist. Soc., 15 (1953) 193–232.

6. Dunham Jackson, Mathematical principles in the theory of small samples, this MONTHLY, 42 (1935) 344–364.

PÓLYA'S ORCHARD PROBLEM

THOMAS TRACY ALLEN

Department of Entomology, 201 Wellman Hall, University of California, Berkeley, CA 94720

1. Introduction. "How thick must the trunks of the trees in a regularly spaced circular orchard grow if they are to block completely the view from the center?" (Pólya and Szegő [**6**].)

Let the trees be circular columns, all of which have the same radius r, so the problem simplifies to one of circles on a plane. Let the circles be centered at integer-valued x and y coordinates, satisfying $(x^2 + y^2)^{1/2} \leqslant s, s$ the radius of the orchard. A *ray* is a straight line of sight from the origin, $(0,0)$. The first circle intersected by a given ray is *visible* along that ray. The problem is to find a radius of the trees, ρ, such that for $r \geqslant \rho$ only trees centered within the orchard are visible, but such that for $r < \rho$ there is a view out along at least one ray. Of course, ρ will be a function of s.

The solution by G. Pólya, based on a method due to A. Speiser (reference cited above, originally in G. Pólya [**5**]), and the solution by R. Honsberger [**4**], based on Minkowskii's Convex Body Theorem, both leave uncertainty about the value of ρ. They show that if s is an integer, then

$$(1) \qquad 1/(s^2 + 1)^{1/2} \leqslant \rho < 1/s.$$

But what is the exact value of ρ? And what is ρ when s is not an integer?

Thomas Tracy Allen: I received my Ph.D. in biophysics from the University of California, Berkeley. A fascination with the intricate dynamics of insect populations led me to the mathematics of coupled nonlinear oscillators, and thence a circuitous path not untypical of science led me to the orchard problem as I present it here. My current vocation is electronic instrumentation, applied to monitoring animal populations and their environments. I have three sons, ages 2, 4, and 8, whose favorite orchard problem is "hide-and-seek".

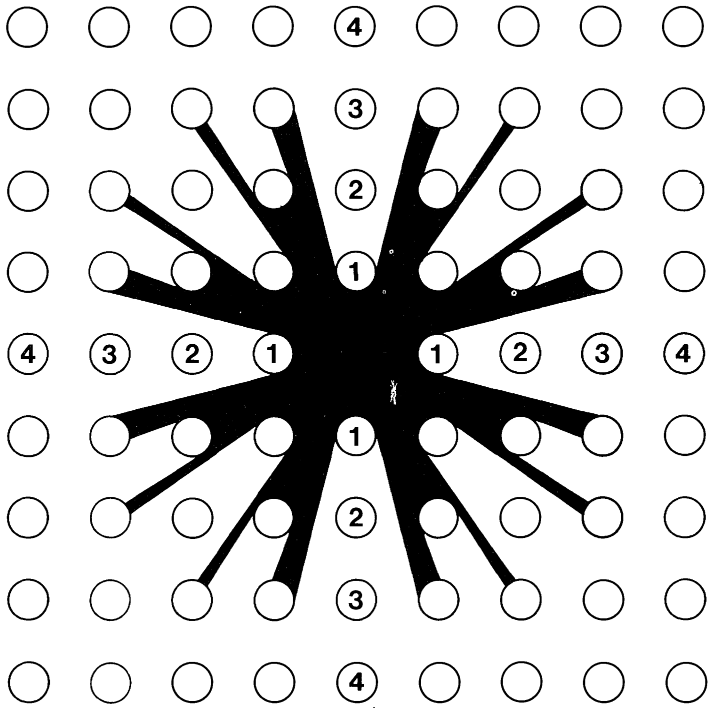


FIG. 1. Rays plotted to their end points on circles, $r \approx .25$.

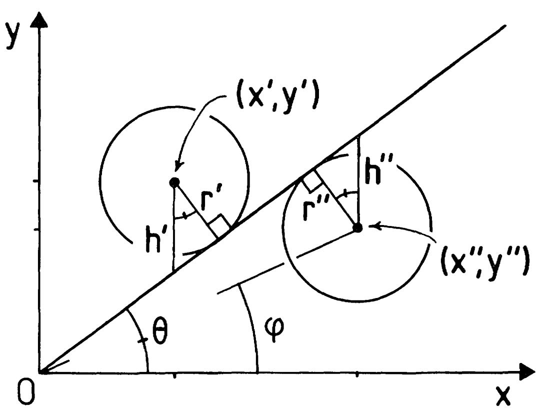


FIG. 2. To calculate the distance r' from the lattice point (x', y') to the ray subtending the angle θ, note that $r' = h' \cos\theta$ and $h' = y' - x' \tan\theta$, so $r' = y'\cos\theta - x'\sin\theta$. If the ray passes through a lattice point, (x, y), then $\sin\theta = y/(y^2 + x^2)^{1/2}$, $\cos\theta = x/(y^2 + x^2)^{1/2}$ and equation (2a) follows—similarly for equation (2b). To calculate θ given r'' and ψ, where $\psi = \arctan y''/x''$, note that $\theta - \psi = \arcsin r''/(x''^2 + y''^2)^{1/2}$. The inequality of (7d) follows directly—similarly for the other inequalities (7a), (7b), and (7c).

Here I show, using elementary methods, that $\rho = 1/\sqrt{\lambda}$, where λ is the first integer greater than s^2 that can be written as a sum of the squares of two coprime integers. If s is itself an integer, then equality on the left hand side of (1) is exact. I also formulate a couple of related problems, to show other aspects of this pretty orchard.

2. Preliminaries. Observe the following: (1) The problem has eightfold symmetry, due to the eight circles nearest the origin. See Fig. 1. It suffices to consider only the first half-quadrant, where

$x \geqslant y \geqslant 0$, $(x, y) \neq (0,0)$. (2) Only circles centered at coprime coordinates can ever be visible. For example, the circle at $(2,2)$ is totally eclipsed by the circle at $(1,1)$, independent of r. (3) In the limit of $r = 0$, all circles—now points—having coprime coordinates are visible; they are first out along a ray. (4) At the other limit, when $r = 1/2$, the circles touch, and only the circles centered at $(1,0)$ and $(1,1)$ (and their images in the four quadrants) are visible.

Consider the ray through (x, y) and the lattice-points lying above and below that ray: (x', y') satisfying $y'/x' > y/x$, and (x'', y'') satisfying $y''/x'' < y/x$. The perpendicular distances from these points to the ray are, in normal form (see Fig. 2),

$$r' = (y'x - x'y)/(x^2 + y^2)^{1/2} \tag{2a}$$

and

$$r'' = (x''y - y''x)/(x^2 + y^2)^{1/2}. \tag{2b}$$

When x, y are coprime (i.e., their greatest common divisor is 1), the moduli in the numerators take on all positive integral values as functions of the (x', y') and the (x'', y''). Therefore, the points closest to the ray are given by

$$y'x - x'y = +1 \tag{3a}$$

and

$$x''y - y''x = +1, \tag{3b}$$

giving a minimum distance of

$$r' = r'' = (x^2 + y^2)^{-1/2}. \tag{4}$$

Equations (3a) and (3b) are indeterminate; they have infinitely many solutions. These are $(kx + x', ky + y')$ and $(kx + x'', ky + y'')$, respectively, where k is any integer, and (x', y') and (x'', y'') are particular solutions. Collectively, these points lie on two lines parallel to the ray and equidistant at the minimum distance given by (4) on either side of it. For visibility, we want the solutions that are closer to the origin than is (x, y), and in the same quadrant, say,

$$0 \leqslant x' < x \tag{5a}$$

and

$$0 \leqslant y'' < y. \tag{5b}$$

There exists exactly one coordinate-pair, (x', y'), satisfying (3a) in the interval given by (5a). To see this, observe that $(kx + x', ky + y')$ places exactly one solution in each half-open interval of length x. Similarly, (x'', y'') is determined uniquely by (3b) in the interval given by (5b).

The above reasoning shows that the circles centered at the two points determined by (3) and (5) and having the radius (4) are tangent to the ray through (x, y). Also, the ray hits both of the points of tangency before it intersects any point of the circle centered at (x, y). To verify this, from the Pythagorean Theorem, the distances from the points of tangency out to (x, y) along the ray are

$$\left[x''^2 + y''^2 - 1/(x^2 + y^2)\right]^{1/2} \quad \text{and} \quad \left[x'^2 + y'^2 - 1/(x^2 + y^2)\right]^{1/2}.$$

The minimum such distance occurs for $(x, y) = (2,1)$, $(x'', y'') = (1,0)$, so the distance must always be greater than or equal to $[1^2 + 0^2 - 1/(2^2 + 1^2)]^{-1/2} = 2/\sqrt{5}$. But from (4), the radius of the circles at tangency is less, namely, less than or equal to $(2^2 + 1^2)^{-1/2} = 1/\sqrt{5}$.

In conclusion, the given circle centered at coprime (x, y) is visible along at least one ray so long as $r < (x^2 + y^2)^{-1/2}$, but it is totally eclipsed when $r \geqslant (x^2 + y^2)^{-1/2}$.

3. The orchard problem. Now we are ready to string a fence at a finite radius, s, around the

observation post at the origin. To start, suppose that the fence is an imaginary circle in an orchard that extends to infinity all around. What is the radius of the trees, $r = \rho$, that just blocks the view of all the trees beyond the fence?

From (4), each tree becomes totally eclipsed when r equals the reciprocal of the distance of its center from the origin. Therefore, the last tree outside the fence to be totally eclipsed as the trees grow must be the one closest to the fence on the outside. All trees farther from the origin were already eclipsed at smaller r. Thus the reciprocal of the distance to the center of the tree closest outside the fence is the desired ρ. Q.E.D.

The reasoning does not depend on having actual trees outside the fence, because the final visibility depends only upon the trees inside.

If s is an integer, then $s^2 + 1$ is the first integer greater than s^2, and there is always a tree centered at $(s,1)$, $(s^2 + 1)^{1/2}$ units from the origin. The critical radius for the view to the outside is therefore $r = \rho = (s^2 + 1)^{-1/2}$. There may be other trees at the same critical distance from the origin. For example, if $s = 50$, then $(50,1)$, $(49,10)$ and 14 images of these two points located in the four quadrants would all disappear at exactly the same radius, $r = \rho = 1/\sqrt{2501}$.

Note that the expression, $\rho = (s^2 + 1)^{-1/2}$, we have found here is the same as that on the left-hand side of (1). The equality there is exact, so long as s is an integer.

If s is not necessarily an integer, a more illuminating way to state the problem is, "What is ρ as a function of s?" Again, let x and y be arbitrary coprime integers. Then $x^2 + y^2 = \lambda$ is, ipso facto, a special kind of integer, one that can be represented as a sum of the squares of two coprime integers in at least one way. Suppose λ_i and λ_{i+1} are two successive integers of this kind, $\lambda_i < \lambda_{i+1}$. Then for $\sqrt{\lambda_i} \leqslant s < \sqrt{\lambda_{i+1}}$, the critical radius ρ must be constant, and it must have the value $\rho = 1/\sqrt{\lambda_{i+1}}$, because $\sqrt{\lambda_{i+1}}$ is the distance to the first center beyond the fence. Or to put it another way independent of s: For $1/\sqrt{\lambda_{i+1}} \leqslant r < 1/\sqrt{\lambda_i}$, the farthest tree visible is centered $\sqrt{\lambda_i}$ units from the origin, because the tree $\sqrt{\lambda_{i+1}}$ units from the origin is already eclipsed when $r = 1/\sqrt{\lambda_{i+1}}$.

To illustrate, in Fig. 3 I have listed pairs of coprime integers in order of increasing $\lambda = x^2 + y^2$. The connecting lines on the figure emphasize that each tree can be eclipsed by exactly two others lying nearer the origin. Now, 65 and 73 are successive λ, interesting because 65 is the first integer that can be written as the sum of the squares of two coprime integers in two different ways. For $1/\sqrt{73} \leqslant r < 1/\sqrt{65}$, the farthest trees visible are the ones at $(8,1)$ and $(7,4)$ (and their images in the four quadrants). Were $r \geqslant 1/\sqrt{65}$ these would disappear simultaneously. And were we to string a fence at any radius in the interval, $\sqrt{61} \leqslant s < \sqrt{65}$, the critical radius for the view out would be $r = \rho = 1/\sqrt{65}$.

4. The continued fraction. Suppose we gaze permanently along a single ray as the trees grow. Which trees will we see?

Let us expand the slope of the ray, $\tan\theta$, into its continued fraction. If p_n/q_n is the nth convergent, then we know that p_n/q_n is a better rational approximation to the slope of the ray than all fractions having denominators less than or equal to q_n. That is, (Hardy and Wright [**3**, Theorem 181]) if $n > 1$, $0 < q \leqslant q_n$, and $p/q \neq p_n/q_n$, then

$$|p_n - q_n \tan\theta| < |p - q\tan\theta|. \tag{6}$$

In geometric terms, (6) says that the vertical distance from (q_n, p_n) to the ray is less than the vertical distances from any of the named (q,p) to the ray. The vertical distances are the h on Fig. 2, including both the h' above and the h'' below the ray. Multiplying both sides of (6) by $\cos\theta$ gives expressions for the perpendicular distances, the r. Since $\cos\theta$ is positive in the first quadrant, the conclusion of the theorem carries over to these perpendicular distances: The point (q_n, p_n) is closer to the ray than are any of the (q,p).

Also we know that $q_n < q_{n+1}$, so this same theorem assures us that (q_{n+1}, p_{n+1}) is closer to the ray than is (q_n, p_n), for all n. Completing the reasoning as before, the circles centered at the

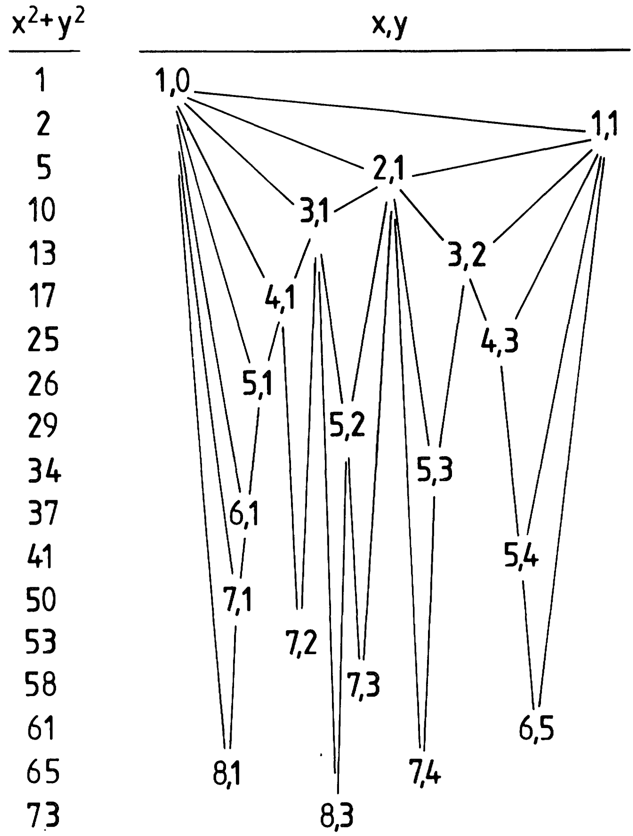


FIG. 3. The points (x, y) are listed in order of increasing $x^2 + y^2$. The connecting lines emphasize that the circle centered at each of these points will be eclipsed by exactly two others centered closer to the origin as the radius increases.

points, $\ldots, (q_{n+1}, p_{n+1}), (q_n, p_n), (q_{n-1}, p_{n-1}), \ldots$ are the ones visible along the ray in sequence as r increases.

The number of convergents and of visible trees is finite if $\tan\theta$ is rational, because we are looking right at a lattice point when $r = 0$. To be correct, the continued fraction must be written with the final coefficient greater than one. Otherwise, as a simple calculation shows, the next-to-the-last and the second-to-the-last convergents give points that are equidistant on opposite sides of the ray. But only the circle closer to the origin is ever visible along that ray. For example, the convergents of $21/16 = [1,3,4,1]$ are $1/1$, $4/3$, $17/13$ and $21/16$. But only the circles at $(1,1)$, $(3,4)$ and $(16,21)$ are ever visible along $\tan\theta = 21/16$. So we write $21/16 = [1,3,5]$.

If $\tan\theta$ is irrational, the number of convergents and of visible trees is infinite. For example, if the slope is $\tan\theta = (\sqrt{5} + 1)/2$, the sequence of circles visible as a function of *decreasing* radius, for $r \leqslant 1/2$, would be the Fibonacci sequence $(1,1), (1,2), (2,3), (3,5), (5,8), (8,13)$, etc. Note that the tree at $(0,1)$ would not be visible along that ray unless the trees overlap, unless $r \geqslant 2\{(\sqrt{5} + 1)^2 + 2^2\}^{-1/2} \approx .53$.

5. Single trees. I claim that the circle centered at arbitrary, coprime (x, y), $x \geqslant y \geqslant 0$, $(x, y) \neq (0,0)$ and for $r \leqslant 1/2$, is visible through the interval of angles (relative to the positive x axis) given by the intersection of the following inequalities:

$$\theta \leqslant \arctan y/x + \arcsin r/\left(x^2 + y^2\right)^{1/2} \quad \text{for } 0 \leqslant r < r^+, \tag{7a}$$

(7b) $$\theta < \arctan y'/x' - \arcsin r/(x'^2 + y'^2)^{1/2}$$
$$\text{for } r^+ \leqslant r < r^0,$$

(7c) $$\theta \geqslant \arctan y/x - \arcsin r/(x^2 + y^2)^{1/2}$$
$$\text{for } 0 \leqslant r < r^-,$$

(7d) $$\theta > \arctan y''/x'' + \arcsin r/(x''^2 + y''^2)^{1/2}$$
$$\text{for } r^- \leqslant r < r^0,$$

where x', y' and x'', y'' satisfy (3) and (5), and where

$r^0 = (x^2 + y^2)^{-1/2}, r^+ = \left[(x + x')^2 + (y + y')^2\right]^{-1/2}$ and $r^- = \left[(x + x'')^2 + (y + y'')^2\right]^{-1/2}$

The reasoning of Section 2 does not establish this result. I need to prove that the circles centered at (x', y') and (x'', y'') determine not only the critical radius, r^0, for total eclipse, but also all degrees of partial eclipse as claimed.

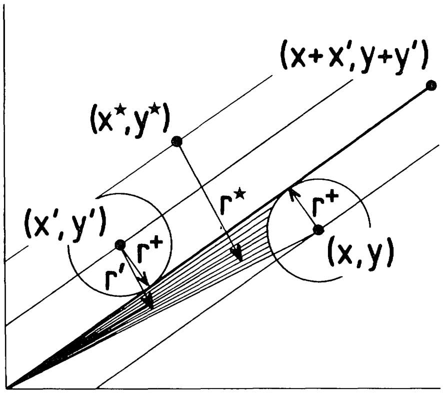


FIG. 4. Circles centered at points (x', y') and (x, y), satisfying (3a) and (5a), are tangent to the ray through $(x + x', y + y')$ when $r = r^+ = \{(x + x')^2 + (y + y')^2\}^{-1/2}$. Above this critical radius, the former circle encroaches from above on the visibility of the latter. The circle centered at (x^*, y^*), $y^*/x^* > y'/x'$, is irrelevant.

Consider the construction shown on Fig. 4. Applying the formulae (2), we find that the points (x, y) and (x', y') are equidistant at the minimum distance r^+ on either side of the ray through $(x + x', y + y')$. Therefore, when $r < r^+$ the upper part of the circle centered at (x, y) is visible and (7a) applies for the upper bound on visibility. (I justify the form of the functions of (x, y) and r in the caption to Fig. 2.) However, when $r^+ \leqslant r < r^0$, the lower ray tangent to the circle centered at (x', y') coincides with rays in the wedge shown shaded on Fig. 4. So the circle centered at (x, y) is partly eclipsed from above. Observe that the point (x', y') is *less than* $2r^+$ distant from all of the rays in the shaded wedge. Other points, (x^*, y^*), $y^*/x^* > y/x, 0 \leqslant x^* < x$, are *more than* $2r^+$ distant from rays in the wedge, because the points of the lattice are (from the modulus in the expression for distance) spaced at multiples of r^+ from the ray through $(x + x', y + y')$. Only (x', y') lies exactly r^+ above the ray in the interval $[0, x)$. Thus (7b) is both necessary and sufficient to describe the upper bound on visibility when $r^+ \leqslant r < r^0$. The same reasoning applies to the effect of the circle centered at (x'', y''), as it eclipses from below, whence (7c) and (7d) for the lower bound on visibility.

On Fig. 5 I have plotted the domains defined by (7a) through (7d) for enough (x, y) to convey the pattern of tessellation of the r-θ plane. Of course, every coprime (x, y) would have an area on that plane, filling it completely as $x, y \to \infty$.

6. Postscript. A problem of visibility in a lattice of triangles (instead of circles) arises from a

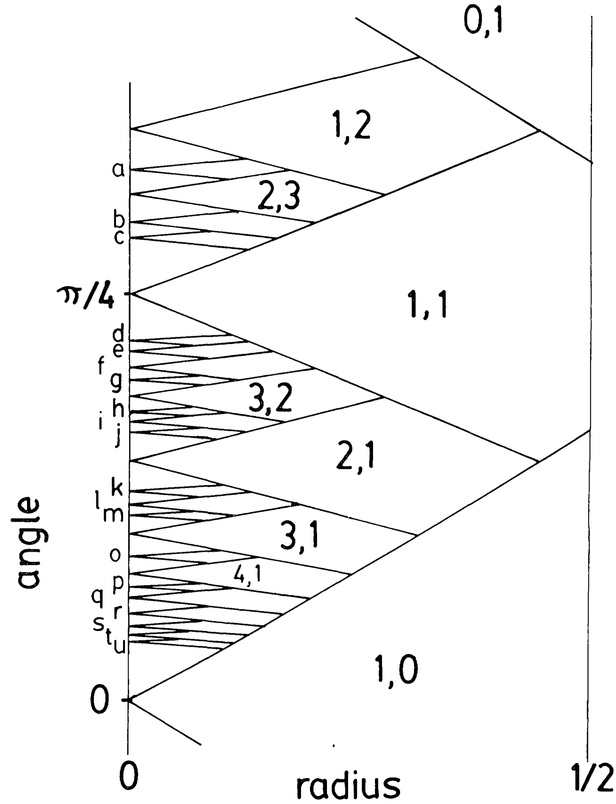


FIG. 5. The intervals of angles through which different circles are visible are functions of the radius of the circles. Letters beside the ordinates correspond to circles centered at: a—(3,5), b—(3,4), c—(4,5), d—(6,5), e—(5,4), f—(4,3), g—(7,5), h—(8,5), i—(5,3), j—(7,4), k—(7,3), l—(5,2), m—(8,3), n—(3,1), o—(7,2), p—(9,2), q—(5,1), r—(6,1), s—(7,1), t—(8,1), u—(9,1).

simple model of coupled neurons, such as the neurons that coordinate the beating of the heart (Allen [**1**]). The same lattice is also a pretty good model of certain electronic circuits, such as those that hold the picture steady on television sets (Allen [**2**]). The solutions are similar to the one given here. For example, a diagram analogous to Fig. 5 depicts how one oscillator can lock onto rational multiples of the period of a second oscillator, depending on the strength of the coupling between them and on the ratio of their natural periods, and indicates how this locking fairs against the destabilizing effects of noise.

Acknowledgements. I thank R. Honsberger for thoughtful suggestions.

References

1. T. T. Allen, On the arithmetic of phase locking: coupled neurons as a lattice on R^2, Phys. D, 6 (1983) 305–320.

2. ______, Complicated, but rational, phase locking responses of a simple time-base oscillator, IEEE Trans. Circuits and Systems, 30 (1983) 627–632.

3. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, Clarendon Press, Oxford, 1979.

4. R. Honsberger, Mathematical Gems I, Dolciani Mathematical Expositions, no. 1, Chap. 4, Mathematical Association of America, Washington, DC, 1973.

5. G. Pólya, Zahlentheoretisches und wahrscheinlichkeits-theoretisches uber die sichtweite im walde, Arch. Math. und Phys., 27, Series 2 (1918) 135–142.

6. G. Pólya and G. Szegő, Problems and Theorems in Analysis, vol. 2, Chap. 5, Problem 239, Springer-Verlag, New York, 1976.

He thought his major contribution was in the calculus of variations, but most of us know him for his real exposition. (See p. 122.)

$x, y \in R$. Thus, since all products are central,

$$xy = (xy)^2 = x(yxy) = (yxy)x = (yx)^2 = yx,$$

and so R is commutative.

THEOREM 2. *If for all* $x, y \in R, (xy - yx)^2 = xy - yx$, *then* R *is commutative.*

Proof. If $yx = 0$, then $xy = (xy)^2 = x(yx)y = 0$, so, by the lemma, $xy - yx \in Z(R)$, for all x and y. Further,

$$xy - yx = (xy - yx)^2 = (yx - xy)^2 = yx - xy,$$

so $x(xy - yx) = (yx - xy)x$. This gives at once that $x^2 \in Z(R)$, for all $x \in R$. Next,

$$\begin{aligned}(xy)^2 = x(yxy) &= x\left[(yx)^2 + y^2 - (yx - y)^2 - y^2x\right] \\ &= \left[(yx)^2 + y^2 - (yx - y)^2 - y^2x\right]x = (yxy)x = (yx)^2,\end{aligned}$$

for all $x, y \in R$. Finally,

$$\begin{aligned}xy - yx = (xy - yx)^2 &= xy(xy - yx) - yx(yx - xy) \\ &= (xy)^2 - xy^2x - (yx)^2 + yx^2y = -x^2y^2 + x^2y^2 = 0,\end{aligned}$$

which means that R is commutative.

REMARKS. One might conjecture the following possible generalizations of Theorem 1 and 2.

(i) Let R be a ring in which $(xyz)^2 = xyz$ for all $x, y, z \in R$. Then R is commutative.
(ii) Let R be a ring in which $(xy)^2 - xy \in Z(R)$ for all $x, y \in R$. Then R is commutative.
(iii) Let R be a ring in which $(xy - yx)^2 - (xy - yx) \in Z(R)$ for all $x, y \in R$. Then R is commutative.

However, the non-commutative ring M of strictly upper triangular 3×3 matrices over Z_3 satisfies the condition $xyz = 0$ for all $x, y, z \in R$ and so is a counterexample to all three conjectures.

In conclusion, we note that Theorem 1 is a special case of the following more general result:

If for each $x, y \in R$, *there exists* $n = n(x, y) > 1$ *such that* $(xy)^n = xy$, *then* R *is commutative.*

Since we are not aware of any readily accessible source of this result in the literature, we indicate here the outline of a proof, which, however, uses relatively heavy machinery.

In a ring R satisfying the given condition, clearly $xy = 0 \Rightarrow yx = 0$. It follows that if x is nilpotent and y is arbitrary, then xy nilpotent. Thus, the nilpotent elements of R annihilate R on both sides, and are therefore central. Since R is obviously periodic, commutativity now follows by a result of Bell [**2**].

References

1. I. N. Herstein, Non-Commutative Rings, Carus Mathematical Monograph No. 15, Mathematical Association of America, Washington, DC, 1968.

2. H. E. Bell, A Commutativity Study for Periodic Rings, Pacific J. Math., 70 (1977) 29–36.

ANSWER TO PHOTO ON PAGE 105

Constantin Carathéodory, 1873–1950, author of many papers on geometric optics and of the immortal *Vorlesungen über reelle Funktionen*.

CYCLIC RELATIVE DIFFERENCE SETS

PHILIP A. LEONARD
Department of Mathematics, Arizona State University, Tempe, AZ 85287

1. Introduction. There is a close connection between difference sets and finite geometries coordinatized by Galois fields. This (with many other items of interest in combinatorics and coding theory) is illustrated most simply by the Fano plane (see Fig. 1). The set $R = \{1,2,4\}$ of quadratic residues modulo 7 is a perfect difference set, since the differences arising from pairs of elements of R give each nonzero residue modulo 7 exactly once. The plane is generated by forming the seven lines $R_j = R + j$, $0 \leqslant j \leqslant 6$, each consisting of three points from the set $S = \{0,1,2,3,4,5,6\}$ of residues modulo 7.

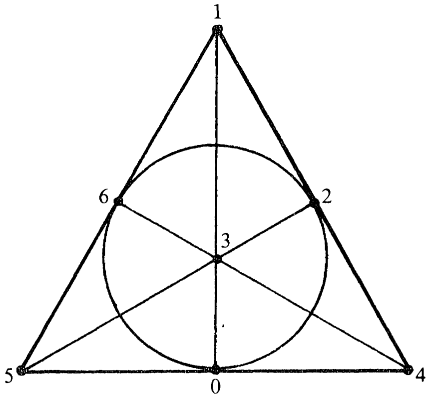


FIG. 1

The connection with Galois fields can be illustrated be means of linear functionals. The field $\mathrm{GF}(2^3)$ can be generated by adjoining to GF(2) an element θ satisfying $\theta^3 + \theta + 1 = 0$. The linear functional $L\colon \mathrm{GF}(2^3) \to \mathrm{GF}(2)$ with $L(1) = 1$, $L(\theta) = L(\theta^2) = 0$ gives rise to the Fano plane, as we have $R = \{i | L(\theta^i) = 0\}$.

More generally, if q is a prime power, $\mathrm{GF}(q^{N+1})$ gives rise to a projective geometry of dimension N coordinatized by $\mathrm{GF}(q)$. That the geometries correspond to difference sets is a famous result due to Singer [**15**]. The first "relative" difference sets were found by Bose [**3**], who proved an affine analogue of the planar version of Singer's theorem. The ideas of Bose underlie §2, in which linear functionals will be used to derive an important family of relative difference sets first studied by Butson [**4**], [**6**]. Most of the applications of cyclic relative difference sets have been to the construction of matrices of combinatorial interest. We conclude this introduction by retracing the well-known path from the Fano plane to a Hadamard matrix of order 8.

The incidence matrix A of the plane ($a_{ij} = 1$ if $j \in R_i$, 0 otherwise, where $(i,j) \in S \times S$) is a 7×7 circulant matrix with first row (0110100). Replacing 0's by -1's, and placing the resulting matrix as the lower right-hand corner of an 8×8 matrix M with remaining entries all 1 gives a Hadamard matrix of order 8 ($MM^t = 8I$). Similar examples will be developed in §3, after a discussion of relative difference sets in §2.

Philip A. Leonard graduated from Boston College and completed his Penn State Ph.D. in 1968 under S. Chowla. He then joined the faculty of Arizona State University, where he is now Professor of Mathematics. Visiting appointments at Carleton University, Ottawa, and the University of Wyoming betray a nostalgia for challenging winters. His mathematical interests center on number theory and aspects of combinatorics. Other pursuits include family camping, something between jogging and running, music in various forms, and Western monastic spirituality.

2. The cyclic relative difference sets of Bose and Butson. A set R of k elements in an Abelian group G of order mn is a *relative difference set* of G relative to a subgroup H of order $n \neq mn$ if, and only if, the differences $r - s$, between distinct elements of R, give only elements of G which are not in H, and give every such element exactly d times.

Such sets will be indicated by $R = R(m, n, k, d)$, and will be called *cyclic* when the group G is cyclic. In this case, G may be taken as the additive group of the integers modulo mn. Ordinary cyclic difference sets arise when $n = 1$.

The set $R = \{0, 1, 4, 9, 11\}$ of residues modulo 24 is an example of a cyclic relative difference set $R(6, 4, 5, 1)$, as each residue which is not a multiple of 6 is represented once as a difference of two elements of R. If the elements of R are reduced modulo 6, the resulting set is an $R(6, 1, 5, 4)$. In general (see [**6**, Theorem 2.1] and also [**2**, Lemma 4.1]), if we start with a cyclic $R(m, n, k, d)$ then, for each factorization $n = st$, we may reduce elements modulo ms to give a cyclic $R(m, s, k, td)$, so a cyclic difference set when $s = 1$.

An extensive discussion of relative difference sets can be found in [**6**]. A more recent paper [**13**] contains additional results, together with a tabulation of cyclic relative difference sets with $k \leqslant 50$. In these two papers, the concept is attributed to Butson [**4**], though the idea appears already in the work of Bose [**3**].

We shall be concerned with cyclic relative difference sets

$$R\big((q^r - 1)/(q - 1), q - 1, q^{r-1}, q^{r-2}\big),$$

for q a prime power, and with sets derived from these in the manner indicated in the previous example using factorizations $q - 1 = st$. The existence of these sets was shown by Butson [**4**], [**6**] using results of Zierler [**20**] on maximal-length linear recurring sequences. They can be derived much more quickly (see below) through the use of linear functionals. (For other treatments, where the flavor of geometry replaces that of linear algebra, see the papers of Berman [**2**] and Jungnickel [**12**] listed in the bibliography.)

For q, a prime power, and $r > 1$, an integer, consider the field $\mathrm{GF}(q^r)$ with q^r elements as a vector space of dimension r over $\mathrm{GF}(q)$. $L: \mathrm{GF}(q^r) \to \mathrm{GF}(q)$ will be an arbitrary nonzero linear functional. A fact which is well known (and easy to establish) is that L gives all nonzero linear functionals. For let θ generate the multiplicative group of $\mathrm{GF}(q^r)$. For $0 \leqslant j \leqslant q^r - 2$ and β in $\mathrm{GF}(q^r)$, define $L_j(\beta) = L(\theta^j\beta)$; the L_j are the $q^r - 1$ distinct nonzero linear functionals in question. This observation prepares the way for the Bose–Butson relative difference sets.

THEOREM. *Let $L: \mathrm{GF}(q^r) \to \mathrm{GF}(q)$ be a nonzero linear functional. Then, considered as a set of residues modulo $q^r - 1$, the exponents i for which $L(\theta^i) = 1$ form a cyclic relative difference set $R((q^r - 1)/(q - 1), q - 1, q^{r-1}, q^{r-2})$.*

Proof. The linear functionals L_j are those described above. For each j, define

$$S_j = \{\beta | L_j(\beta) = 1\} \quad \text{and} \quad R_j = \{i | \theta^i \in S_j\}.$$

Each R_j consists of $k = q^{r-1}$ residues modulo $q^r - 1$; moreover, if

$$R_0 = \{d_1, \dots, d_k\}$$

then, as $1 = L(\theta^d) = L(\theta^j \cdot \theta^{d-j})$, we have

$$R_j = \{d_1 - j, \dots, d_k - j\}.$$

Thus elements of $R_0 \cap R_j$ are in one-to-one correspondence with pairs (d, d') of elements of R having $j \equiv d - d' (\mathrm{mod}\, q^r - 1)$. In order to show that R is a relative difference set as claimed, we must consider the intersections $R_0 \cap R_j$.

Case 1. $j \equiv 0$ (mod $(q^r - 1)/(q - 1)$).

In this case θ^j lies in $\mathrm{GF}(q)$. The linearity of L and $L(\beta) = 1$ thus imply $L_j(\beta) = \theta^j \neq 1$, so

that $R_0 \cap R_j$ is empty. Thus j does not occur as a difference of elements from $R = R_0$.

Case 2. $j \not\equiv 0 \pmod{(q^r - 1)/(q - 1)}$.

In this case the cosets S_0 and S_j are cosets of distinct subspaces of rank $r - 1$ in $\mathrm{GF}(q^r)$. The underlying subspaces are $\ker L$ and $\ker L_j$, and these intersect in q^{r-2} points since

$$\dim(T + U) = \dim T + \dim U - \dim(T \cap U)$$

for all subspaces T, U. Because $\theta^j - 1$ does not lie in $\mathrm{GF}(q)$, β in $\ker L$ can be chosen with $L(\gamma) \neq 0$, where $\gamma = (\theta^j - 1)^{-1}\beta$. But then

$$L(\theta^j\gamma) = L(\gamma) = c \neq 0,$$

and $\alpha = c^{-1}\gamma$ satisfies $L(\alpha) = L_j(\alpha) = 1$, so that α lies in $S_0 \cap S_j$. Translating this intersection by $-\alpha$ preserves cardinality, and gives $\ker L \cap \ker L_j$. Thus $S_0 \cap S_j$ contains q^{r-2} points, as required.

Our example of $R = \{0,1,4,9,11\}$ modulo 24 is obtained easily from the theorem just established. Consider the generation of $\mathrm{GF}(5^2)$ over $\mathrm{GF}(5)$ by the adjunction of a root θ of the primitive polynomial $X^2 + 2X + 3$. If L is defined by setting $L(1) = L(\theta) = 1$ and extending linearly, then $L(\theta^i) = 1$ exactly for $i = 0,1,4,9,11$.

As noted earlier, a cyclic $R(m,n,k,d)$ gives a cyclic $R(m,s,k,td)$ when $n = st$. Taking $s = 1, t = q - 1$ in the Theorem, we obtain

$$R\big((q^r - 1)/(q - 1), 1, q^{r-1}, q^{r-2}(q - 1)\big).$$

This ordinary difference set is (the complement of) a Singer set. (See [**1**], where r is replaced by $N + 1$.) A second example will further illustrate this idea.

If θ is a root of the primitive cubic polynomial $X^3 + 2X + 1$ over $\mathrm{GF}(3)$ and $L\colon \mathrm{GF}(3^3) \to \mathrm{GF}(3)$ satisfies $L(1) = L(\theta) = L(\theta^2) = 1$, the above construction gives the set

$$R = \{0,1,2,8,11,18,20,22,23\} \text{ modulo } 26.$$

Reducing this modulo 13 gives $\{0,1,2,5,7,8,9,10,11\}$. The complementary set $\{3,4,6,12\}$ is a difference set which can be used to generate a finite projective plane having thirteen points and thirteen lines. (In fact, if the linear functional L is replaced by L_3 from the proof of the theorem, we obtain the Singer difference set $\{0,1,3,9\}$ given by Ryser [**18**, p. 132] for this plane.)

3. Matrix applications of Bose-Butson relative difference sets. The family of cyclic relative difference sets developed in §2 was introduced by Butson as a means of obtaining generalized Hadamard matrices. Other authors have made similar applications of these sets. In this section, I want to explain how these sets lead to the construction of some interesting matrices. The object is to give a simplified derivation of some results of Berman [**2**], and to provide an introduction to some other results in which the Bose-Butson relative difference sets have played a part. It should be pointed out that, for the constructions which follow, any relative difference sets with the indicated parameters will do. The exact nature, or method of construction, of these sets is immaterial.

Berman [**2**] discussed the Bose-Butson relative difference sets in connection with *generalized weighing matrices* $W(d,k,m)$. These are $m \times m$ matrices A whose nonzero entries are dth roots of unity and which satisfy $AA^* = kI$, so that there are k nonzero entries per row. When $m = k$, these are the generalized Hadamard matrices of Butson [**4**]; when $d = 2$ they are "weighing matrices" (see [**7**]), with important special cases being $W(2, m-1, m)$ (Conference matrices [**5**], [**8**]) and $W(2,k,k)$ (Hadamard matrices). I shall sketch the derivation of some of these matrices from the Bose-Butson relative difference sets by means of the Hall polynomials $\Sigma_{r \in R} x^r$ attached to a cyclic relative difference set R. These polynomials are important in the study of cyclic difference sets [**1**], [**9**]. Those arising from Bose-Butson cyclic relative difference sets have been used in connection with conference matrices [**5**] and with Hadamard matrices [**16**]–[**18**]. A treatment of the Singer difference sets along the lines followed here has also been given [**10**].

Because we are concerned with relative difference sets modulo mn, we more properly consider the polynomial

$$\phi(x) = \sum_{r \in R} x^r (\bmod x^{mn} - 1).$$

In what follows we also use the polynomials

$$T_j(x) = 1 + x + \cdots + x^{j-1} = (x^j - 1)/(x - 1), \quad j > 1.$$

The assertion that $R = R(m, n, k, d)$ is a relative difference set translates into the polynomial congruence

$$(*) \qquad \begin{aligned} \phi(x)\phi(x^{-1}) &\equiv k + d\{T_{mn}(x) - T_n(x^m)\} \\ &\equiv k + dT_n(x^m)\{T_m(x) - 1\}(\bmod x^{mn} - 1). \end{aligned}$$

We shall be interested in forming $m \times m$ matrices with entries either zero or nth roots of unity. For this reason (see below) we shall interpret the congruences just given modulo $x^m - \omega$, where $\omega^n = 1$.

The connection with matrices is easy to make. If $P = P(\omega)$ is the $m \times m$ matrix having ω in the lower left-hand corner, 1's on the first superdiagonal, and 0's elsewhere, then $P^m = \omega I$, where I is the $m \times m$ identity matrix, and the matrices $a_0 I + \cdots + a_{m-1}P^{m-1}$ form an algebra isomorphic to that formed by the polynomials

$$a_0 + a_1 x + \cdots + a_{m-1}x^{m-1} \text{ modulo } x^m - \omega;$$

thus statements about polynomials have matrix counterparts. Finally, the matrix

$$a_0 I + a_1 P + \cdots + a_{m-1}P^{m-1}$$

is *ω-circulant*; its first row is

$$(a_0, a_1, a_2, \ldots, a_{m-1}),$$

its second is

$$(\omega a_{m-1}, a_0, a_1, \ldots, a_{m-2}),$$

its third is

$$(\omega a_{m-2}, \omega a_{m-1}, a_0, \ldots, a_{m-3}),$$

and so on. (When $\omega = 1$, such matrices are called *cyclic* or *circulant* matrices; when $\omega = -1$, they are called *negacyclic*.)

If A is an ω-circulant matrix with first row $(a_0, a_1, \ldots, a_{m-1})$, then its conjugate transpose $A^* = B$ is also ω-circulant. Its first row $(b_0, \ldots, b_{m-1})$ has $b_0 = \bar{a}_0$, and $b_j = \bar{\omega} a_{m-j}$ for $j = 1, 2, \ldots, m-1$. Of interest here is the case when A arises from $R = R(m, n, k, d)$ via the Hall polynomial $\phi(x) = \Sigma_{r \in R} x^r$ reduced modulo $x^m - \omega$. The term x^r, for $r = mu + v$, $0 \leqslant v \leqslant m - 1$, reduces to $\omega^u x^v$. The corresponding term x^{-r} in $\phi(x^{-1})$ reduces to $\bar{\omega}^{u+1}x^{m-v}$. From this it follows that, if we arrange (by shifting) that 0 belongs to R, the ω-circulant matrix arising when $\phi(x^{-1})$ is reduced modulo $x^m - \omega$ is exactly A^*. Reduced modulo $x^m - \omega$, the congruence (*) gives

$$(**) \qquad \phi(x)\phi(x^{-1}) \equiv \begin{cases} k(\bmod x^m - \omega), & \text{if } \omega \neq 1, \\ k + dn\{T_m(x) - 1\}(\bmod x^m - 1), \end{cases}$$

since, if $\omega^n = 1, \omega \neq 1$, then $T_n(\omega^m) = 0$. By means of the matrix isomorphism, we thus have an $m \times m$ matrix A which is ω-circulant, and which satisfies

$$AA^* = \begin{cases} kI, \omega \neq 1, \\ kI + dn(J - I), \omega = 1, \end{cases}$$

where J is the matrix all of whose entries are 1. For example, the set $R(6,4,5,1) = \{0,1,4,9,11\}$ modulo 24 gives the 6×6 i-circulant matrix A with first row $(1,1,0,i,1,i)$, satisfying $AA^* = 5I$.

4. "Balance" and a final example. The matrices introduced here enjoy the property of being *balanced*. For a generalized weighing matrix A, this means that the matrix M formed from A by replacing each nonzero entry by 1 satisfies the matrix equation

$$MM^T = rI + \lambda(J - I)$$

for integers r and λ. (Thus, as M would already satisfy $MJ = kJ$, M is the incidence matrix of a symmetric balanced incomplete block design. See [**9**]. For A's restricted to entries $0, 1, -1$, see [**7**] for an extensive discussion.) If A is one of the ω-circulant matrices discussed above, then M is obtained in the same manner by reducing $\phi(x)$ modulo $x^m - 1$ instead of modulo $x^m - \omega$. When this is done $\phi(x)$ becomes the Hall polynomial of the cyclic difference set $R(m,1,k,nd)$ to which R reduces, and

$$MM^T = kI + nd(J - I)$$

from what is noted above for $\omega = 1$.

This observation leads to an interesting example of a balanced weighing matrix. If θ is chosen to satisfy $\theta^4 = \theta^3 + 1$ over GF(3) and the linear functional $L\colon \mathrm{GF}(3^4) \to \mathrm{GF}(3)$ has $L(\theta^j) = 1$ for $j = 0,1,2,3$, the relative difference set

$$\begin{aligned} R &= R(40,2,27,9) \\ &= \{0,1,2,3,6,8,9,11,18,19,23,24,27,33,44,47,50,52,54,56,57,61,68,69,70,72,76\} \end{aligned}$$

is obtained. The construction outlined above leads very easily to a negacyclic balanced weighing matrix of order 40. The first row is given in abbreviated form as

$$(++++-0+\ -++-+-0-0-\ -++0-0+\ +00+\ ---0-\ +00-000).$$

Note that this is not a group matrix in the sense of Mullin and Stanton [**13**], nor is it related to an (α,β) decomposition of a difference set in the sense of Jeffcott and Spears, as there is no such decomposition for a $(40,27,18)$ cyclic difference set [**11**].

5. Acknowledgements. The author is grateful to the Department of Mathematics at the University of Wyoming for support and mathematical stimulation during a sabbatical visit, and to John F. Dillon for conversations related to the subject of this note.

References

1. L. Baumert, Cyclic Difference Sets, Springer Lecture Notes 182, Berlin-Heidelberg-New York, 1971.

2. G. Berman, Families of generalized weighing matrices, Canad. J. Math., 30 (1978) 1016–1028.

3. R. C. Bose, An affine analogue of Singer's theorem, J. Indian Math. Soc., 6 (1942) 1–15.

4. A. T. Butson, Relations among generalized Hadamard matrices, relative difference sets, and maximal length linear recurring sequences, Canad. J. Math., 15 (1963) 42–48.

5. P. Delsarte, J. M. Goethals, and J. J. Seidel, Orthogonal matrices with zero diagonal, II, Canad. J. Math., 23 (1971) 816–832.

6. J. E. H. Elliot and A. T. Butson, Relative difference sets, Illinois J. Math., 10 (1966) 517–531.

7. A. V. Geramita and J. Seberry, Orthogonal Designs, Marcel Dekker, New York, 1979.

8. J. M. Goethals and J. J. Seidel, Orthogonal matrices with zero diagonal, Canad. J. Math., 19 (1967) 1001–1010.

9. M. Hall, Jr., Combinatorial Theory, Blaisdell, Waltham, 1967.

10. H. T. Hsu and J. E. Olson, Application of coding theory to the construction of the Singer difference sets, Int. J. Control, 17 (1973) 1247–1250.

11. B. Jeffcott and W. T. Spears, Decompositions of difference sets: a generalization of balanced weighing matrices, Proc. Eighth Southeastern Conf. on Combinatorics, Graph Theory and Computing (Baton Rouge, 1977), Congressus Numerantium, 19 (1977) 385–414.

12. D. Jungnickel, On automorphism groups of divisible designs, Canad. J. Math., 34 (1982) 257–297.

13. C. W. H. Lam, On relative difference sets, Proc. Seventh Manitoba Conf. on Numerical Math. and Computing (Winnipeg, 1977), Congressus Numerantium, 20 (1977) 445–474.

14. R. C. Mullin and R. G. Stanton, Group matrices and balanced weighing designs, Utilitas Math., 8 (1975) 277–301.

15. J. Singer, A theorem in finite geometry and some applications to number theory, Trans. Amer. Math. Soc., 43 (1938) 377–385.

16. E. Spence, Hadamard matrices from relative difference sets, J. Combin. Theory Ser. A, 19 (1975) 287–300.

17. ______, Skew-Hadamard matrices of order $2(q+1)$, Discrete Math., 18 (1977) 79–85.

18. ______, Skew-Hadamard matrices of the Goethals-Seidel type, Canad. J. Math, 27 (1975) 555–560.

19. H. J. Ryser, Combinatorial Mathematics, Carus Mathematical Monographs, no. 14, Mathematical Association of America, Washington, DC, 1963.

20. N. Zierler, Linear recurring sequences, SIAM J. Appl. Math., 7 (1959) 31–48.

ON "COS $F(X) = F$(SIN X)"

T. L. McCoy

Department of Mathematics, Michigan State University, East Lansing, MI 48824-1027

This note is concerned with the question: do there exist differentiable functions f satisfying the functional equation

$$\cos f(x) = f(\sin x)? \tag{1}$$

This question, a variant of a problem appearing in the MONTHLY (E 2478) was one of the problems appearing on the Herzog Prize Examination, a yearly undergraduate mathematics competition at Michigan State University. Several contestants answered the question affirmatively by noticing that (1) has a constant solution, but clearly it would be more satisfying to know whether (1) has other, less trivial, solutions. We will see that (1) has a family of non-constant solutions which are C^∞ on $\mathbb{R}$ (but, it turns out, none which are analytic at the origin).

The equation above is of the form

$$\phi(f(x)) = \psi(x, f(x)) \tag{2}$$

concerning which, the author has recently learned, there is a substantial literature. Equation (2) is studied in great generality in [**3**]. An interesting recent related paper is [**2**], in which sufficient conditions are found for (2) to have a unique continuous solution for (the unknown function) f, an interesting contrast to what happens in our case. It appears that consideration of equations of the form (2) almost inevitably leads to the study of orbits under functional iterates, a technique common to the present paper and the references cited.

Acknowledgment. The author wishes to thank Professor Charles MacCluer for suggesting this problem, and wishes to note that the second of the proofs of non-analyticity below was pointed out by him.

The constant solution alluded to in our second paragraph is $f(x) = \tau$, where τ satisfies

$$\cos \tau = \tau. \tag{3}$$

It is easy to show (3) defines τ uniquely. In this note τ will always have this meaning. Notice that any solution of (1) which is defined at $x = 0$ must satisfy $f(0) = \tau$.

The author received his B.A. from Oberlin College in 1954 and a Ph.D. in Mathematics from the University of Wisconsin in 1961, working in the area of complex analysis under the direction of J. Korevaar. He taught at Illinois Institute of Technology from 1960 to 1964, and at Michigan State University from 1964 to the present, and has spent sabbatical years at Glasgow University and the University of New Mexico. Some non-mathematical interests and activities are travel, swimming and hiking, and music both as listener and as pianist. A favorite pastime is joining other amateur musicians to play sonatas and chamber music. The author has two children, a daughter planning to attend MSU and a son already embarked on a business career.

We shall make much use of a certain partition of $(0, \pi/2)$ into subintervals. Set $x_0 = \pi/2$ and $x_{n+1} = \sin x_n$ for $n \geqslant 0$, and set $I_n = [x_n, x_{n-1}]$. Clearly the I_n are abutting intervals with disjoint interiors. Furthermore, their union fills $(0, \pi/2]$. To see this, note that $\{x_n\}$ is a decreasing sequence and hence has a limit, call it X, and X satisfies $X = \sin X$ proving $X = 0$.

LEMMA 1. *Let f be a solution of* (1) *on* $(0, \pi/2]$ *such that for some* N, $f(I_N) \subset [\tau/2, 3\tau/2]$. Then

$$- \sin\tau < \frac{f(\sin x) - \tau}{f(x) - \tau} \leqslant 0, \quad x \in I_n,\ n \geqslant N. \tag{4}$$

Proof. Let $x \in I_N$. Since $\tau < 1$ we have $3\tau/2 < \pi/2$, so $f(x)$ lies in the principal value range of arc cos; therefore (1) yields

$$\begin{aligned} f(x) &= \operatorname{Arc}\cos f(\sin x) \\ &= \pi/2 - \int_0^{f(\sin x)} \frac{dt}{\sqrt{1-t^2}}. \end{aligned}$$

Since $f(0) = \tau$ we have

$$\begin{aligned} f(x) - \tau &= f(x) - f(0) \\ &= -\int_{f(0)}^{f(\sin x)} \frac{dt}{\sqrt{1-t^2}} \\ &= -\int_{\tau}^{f(\sin x)} \frac{dt}{\sqrt{1-t^2}} \\ &= -\frac{f(\sin x) - \tau}{\sqrt{1-T^2}}, \end{aligned} \tag{5}$$

for some T between τ and $f(\sin x)$. Now $f(\sin x) = \cos f(x)$ lies between $\cos\tau/2$ and $\cos 3\tau/2 < \cos\tau$, thus $\sqrt{1 - T^2}$ lies between $\sin\tau/2$ and $\sin\tau$.

This proves (4) for $n = N$, and also shows f satisfies the hypotheses of the lemma when I_N is replaced by I_{N+1}; thus we can iterate the above argument to obtain (4) for all $n \geqslant N$. □

LEMMA 2. *Let f be a solution of* (1) *on* $[0, \pi/2]$ *such that* $f(I_N) \subset [\tau/2, 3\tau/2]$. *Then* $f_R'(0)$ *exists, and* $f_R'(0) = 0$.

Proof. Iterating (4) yields

$$|f(x) - \tau| < A \sin^m\tau, \quad x \in I_{N+m}, \tag{6}$$

with

$$A = \max_{I_N} |f(x) - \tau|.$$

For any $\sigma < 1$ we can choose N so large that

$$\sin x_n / x_n \geqslant \sigma, \quad n \geqslant N.$$

Then for $x \in I_{N+m}$ we have

$$x \geqslant x_{N+m} \geqslant \sigma x_{N+m-1} \geqslant \cdots \geqslant \sigma^m x_{N-1}.$$

Putting this inequality together with (6) we have

$$\left| \frac{f(x) - f(0)}{x} \right| < \frac{A}{x_{N-1}} \left(\frac{\sin\tau}{\sigma} \right)^m, \quad x \in I_{N+m}. \tag{7}$$

We can suppose $\sigma > \sin\tau$. Thus letting $m \to \infty$ in (7) proves the lemma. □

We now construct a non-constant function f which will eventually be shown to be in

$C^\infty[0, \pi/2]$. We define f arbitrarily on I_0 save for the four requirements

(i) $f(I_0) \subset [\tau/2, 3\tau/2]$,

(ii) $f(x_0) = f(x_1) = \tau$,

(iii) $f \in C^\infty$ on I_0,

(iv) $f_L^{(k)}(x_0) = f_R^{(k)}(x_1) = 0, \quad k \geqslant 1$.

(Here, as is customary with functions defined on closed intervals, we understand continuity at endpoints in terms of one-sided limits).

Now use (1) to extend f to I_1. Because of condition (1) we can apply Lemma 1 to conclude $f(I_1) \subset [\tau/2, 3\tau/2]$. Condition (ii) ensures f is continuous on $I_0 \cup I_1$, and also that $f(x_2) = \tau$. Differentiation of (1) gives

$$-f'(x)\sin f(x) = \cos x \cdot f'(\sin x), \tag{8}$$

and one proves by induction that

$$\begin{aligned} -f^{(k)}(x)\sin f(x) &+ \{\text{terms involving } f^{(j)}(x), \quad j < k \text{ only}\} \\ &= \cos^k x \cdot f^{(k)}(\sin x) + \{\text{terms involving } f^{(j)}(\sin x), \quad j < k \text{ only}\}. \end{aligned} \tag{9}$$

From (8) and (9) we see that f is C^∞ on $[x_2, x_1)$, and because of (iv) f is in fact C^∞ on $I_1 \cup I_0$ since $f_L^{(k)}(x_1) = f_R^{(k)}(x_1) = 0$ (as follows easily from (9) and induction on k). Furthermore, f satisfies the above conditions (i)–(iv) relative to interval I_1, so we are set up to extend f to I_2, then I_3, and eventually to $(0, \pi/2]$. Setting $f(0) = \tau$, we have f defined on $[0, \pi/2]$ and C^∞ on $(0, \pi/2]$. It remains to prove $f_R^{(k)}(0)$ exists for all k, which will not be difficult after the following lemma:

LEMMA 3. *Let f be as in the preceding two paragraphs. Then for each $k \geqslant 1$ there exist C_k and A_k, $0 \leqslant A_k < 1$, such that*

$$|f^{(k)}(x)| < C_k A_k^n, \quad x \in I_n. \tag{10}$$

Proof. From (8),

$$-f'(\sin x)/f'(x) = \frac{\sin f(x)}{\cos x}, \tag{11}$$

and since Lemma 2 implies f is continuous from the right at $x = 0$, we find from (11) that

$$-f'(\sin x)/f'(x) \sim \sin\tau, \quad x \to 0.$$

Therefore we may choose N so large that

$$|f'(\sin x)/f'(x)| < \sin^{1/2}\tau, \quad x \in I_n, \quad n \geqslant N,$$

and iterating this gives

$$|f'(x)| < \sin^{m/2}\tau \cdot \max_{u \in I_N}|f'(u)|, \quad x \in I_{N+m}.$$

This establishes Lemma 3 for $k = 1$, with $A_1 = \sqrt{\sin\tau}$ and $C_1 = \max_{u \in I_N}|f'(u)|$.

Now assume inductively that (10) holds for $j \leqslant k - 1$. Let $P(x)$, $Q(x)$ be the brackets on the left-hand and right-hand sides, respectively, of (9). Both P and Q are polynomials in their dependence on $f^{(j)}(x)$ and $f^{(j)}(\sin x)$, the coefficients being bounded functions of x. We may assume the A_j are an increasing sequence for $0 \leqslant j \leqslant k - 1$; thus for some constant C_k' we have

$$|P(x) - Q(x)| < C_k' A_{k-1}^n, \quad x \in I_n. \tag{12}$$

From (9) we get

$$f^{(k)}(\sin x) = -\sec^k x \cdot f^{(k)}(x) \cdot \sin f(x) + \sec^k x \cdot [P(x) - Q(x)]. \tag{13}$$

If we choose B_k so that $A_{k-1} < B_k < 1$, then (12) yields

$$|P(x) - Q(x)| < B_k^n, \quad x \in I_n, \quad n \text{ sufficiently large}.$$

We can also assume by taking n sufficiently large (which means taking x sufficiently small) that

$$|\sin f(x)| < \sqrt{\sin\tau}$$

and

$$\sec^k x < \csc^{1/4}\tau.$$

Then (13) yields

$$|f^{(k)}(\sin x)| < \sin^{1/4}\tau \cdot |f^{(k)}(x)| + \csc^{1/4}\tau \cdot B_k^N, \quad x \in I_N, \tag{14}$$

for some sufficiently large N. Iterating the above gives us

$$\begin{aligned}|f^{(k)}(\sin\sin x)| &< \sin^{1/4}\tau \cdot |f^{(k)}(\sin x)| + \csc^{1/4}\tau \cdot B_k^{N+1}\\ &< \sin^{1/2}\tau \cdot |f^{(k)}(x)| + B_k^N + \csc^{1/4}\tau \cdot B_k^{N+1}, \quad x \in I_N.\end{aligned}$$

Let us define the sequence $\{S_n(x)\}$ of functions by

$$S_1(x) = \sin x, \quad S_m(x) = \sin S_{m-1}(x).$$

Then iterations of (14) yield

$$|f^{(k)}(S_m(x))| < \sin^{m/4}\tau \cdot |f^{(k)}(x)| + \csc^{1/4}\tau \cdot \sum_{v=0}^{m-1} \sin^{v/4}\tau \cdot B_k^{N+m-1-v}, \quad x \in I_N,$$

whence

$$|f^{(k)}(S_m(x))| \leqslant \sin^{m/4}\tau \cdot |f^{(k)}(x) + m\csc^{1/4}\tau \cdot B_k^N \max_{0\leqslant v\leqslant m-1} \sin^{v/4}\tau \cdot B_k^{m-v}, \quad x \in I_N. \tag{15}$$

It is a fact, easily verified, that if $0 < a < b$, then $a^t b^{m-t}$ is a decreasing function of t. We increase B_k if necessary so that $B_k > \sin^{1/4}\tau$, thus we will have

$$\max_{0\leqslant v\leqslant m-1} \sin^{v/4}\tau \cdot B_k^{m-v} = B_k^m;$$

thus (15) yields

$$\begin{aligned}|f^{(k)}(S_m(x))| &\leqslant \sin^{m/4}\tau \cdot |f^{(k)}(x)| + m\csc^{1/4}\tau \cdot B_k^N B_k^m\\ &:= c_1\sin^{m/4}\tau + c_2 B_k^m\\ &< (c_1 + c_2)\left(\operatorname{Max}(B_k, \sin^{1/4}\tau)\right)^m,\end{aligned}$$

$x \in I_N$ with N sufficiently large, and this clearly implies an estimate of the form (10), where we take $A_k < 1$, but larger than B_k or $\sin^{1/4}\tau$. □

LEMMA 4. *The function f of the preceding paragraph has right-hand derivatives of all orders at the origin* (*and thus is a C^∞ solution of* (1) *on* $[0, \pi/2]$).

Proof. Assume inductively that $f_R^{(j)}(0)$ exists for $j \leqslant k$, and has the value 0. We shall prove that $f_R^{(k+1)}(0) = 0$. We note that Lemma 2 provides a basis for our induction.

We have

$$\left|\frac{f^{(k)}(x) - f^{(k)}(0)}{x}\right| = \left|\frac{f^k(x)}{x}\right| < C_k A_k^{N+m} / x_{N+m}, \quad x \in I_{N+m}, \tag{16}$$

by Lemma 3. The number N is at our disposal; let it be so large that

$$\sin x_n / x_n \geqslant \sigma > A_k, \quad n \geqslant N.$$

An iterative argument used earlier gives

$$x_{N+m} \geqslant \sigma^m / x_{N-1},$$

so (16) implies

$$\left|\frac{f^{(k)}(x)}{x}\right| < C_k \frac{A_k^N}{x_{N-1}}\left(\frac{A_k}{\sigma}\right)^m, \quad x \in I_{N+m}.$$

Now fix N and let $m \to \infty$ to conclude that $f^{(k+1)}(0)$ exists and has the value 0. □

COROLLARY. *There is a non-constant solution to* (1) *which is in* $C^\infty(\mathbb{R})$.

Proof. The solution on $[0, \pi/2]$ defined above extends to $C^\infty[-\pi/2, \pi/2]$ by requiring $f(x) - \tau$ to be even, and can then be extended to $\mathbb{R}$ by the equation $f(x + \pi) = f(x)$. It is clear that this extension is in $C^\infty(\mathbb{R})$, and a little thought shows us it is still even. We check the functional equation (1) as follows:

$$f(\sin(x \pm \pi)) = f(-\sin x) = f(\sin x) = \cos f(x) = \cos f(x \pm \pi).$$

This shows (1) is valid on $[-\pi/2, 0]$. Also, validity of (1) on an interval of length π is seen to imply validity of (1) on either abutting interval of length π, so the region of validity of (1) in fact extends to $\mathbb{R}$. □

As a final comment, we point out that it is easy to see that the only solution of (1) which is analytic at the origin is the constant solution. In fact, if f is any solution which is infinitely differentiable at the origin, then (8) and (9) show $f^{(k)}(0) = 0$, for all $k > 0$, so a power series about the origin would reduce to a constant. A second proof can be based on (4), which shows $f(x) - \tau$ changes sign infinitely often in any interval $[-\varepsilon, \varepsilon]$, hence f' has an infinite sequence of zeros tending to the origin.

References

1. Problem E 2478, this MONTHLY, 81 (1974) 516.

2. Von Manfred Krüppel, Ein Eindeutigkeitssatz für stetige Lösungen von Funktional gleichungen, Publ. Math. Debrecen, 27 (1980) no. 3-4, 201–205.

3. Marek Kuczma, Functional Equations in a Single Variable, PWN–Polish Scientific Publishers, Warszawa, 1968.

---

161. **MISCELLANEA**

Elegy

What has become of the rule of three,
  Simple or double, once popular pair?
Students today no longer see
  Alligation, or tret and tare.
Horner's method, that used to fill
  Many an hour, is now passé.
Over these losses, whose tears will spill?
  That was the math of yesterday.

Instructor, ponder this codicil,
  The awkward truth that you can't gainsay:
What you're teaching now, with so much good will,
  Is tomorrow's math of yesterday.

R. P. Boas
Department of Mathematics
Northwestern University
Evanston, IL 60201

DESIGNING BUFFON'S NEEDLE FOR A GIVEN CROSSING DISTRIBUTION

JACK M. ROBERTSON
Department of Mathematics, Washington State University

ANDREW F. SIEGEL
Department of Statistics and Department of Finance, Business Economics, and Quantitative Methods, University of Washington, Seattle

1. Introduction. In 1733, Georges Louis Laclerc, Comte de Buffon (1707–1788), submitted a memoire to the Académie des Sciences in which he considered a number of problems, including the needle problem, which he solved with geometric probability arguments. This famous problem involves the computation of the probability that a needle thrown will cross a line of a system of evenly spaced parallel lines. This work was not published until 1777 when it appeared in his *Supplement to Natural History*. The breadth of Buffon's work contained in the 36 volumes of his *Natural History* has led to acclaim from diverse sources. The work dealt with such varied topics as probability, astronomy (cooling of planets), physics, plant physics (tensile strength of wood for boat design), forestry, physiology, pyrotechnics (aerial rockets), zoology, botany, and mineralogy [**8**].

As would be expected for such an important method, variations on the theme of Buffon's original needle problem have been many. It is clear he envisaged a quite general class of problems himself. "But if, instead of throwing into the air a round object such as a coin, one threw an object of another shape, such as a square Spanish pistole, or a needle, a stick, etc., the problem would demand a little more geometry, although in general it would always be possible to give its solution by comparisons of space, as we are going to demonstrate" [**2**]. The needle has been lengthened and bent [**4**], [**5**], [**7**]; intersection probabilities between more general families of geometric families of geometric figures have been described [**1**], [**10**]; the grid has been modified so as to improve statistics which approximate π [**6**], [**9**], and partial answers given to the inverse problem in which a curved needle is to be constructed so as to generate a given probability distribution for the number of crossings [**3**].

Let the plane be covered with a grid of parallel lines one unit apart and assume a curve $\mathscr{C}$ of finite length $L(\mathscr{C})$ is randomly cast onto the parallel grid in the sense that the distance from a given point on $\mathscr{C}$ to the nearest line in the grid is uniformly distributed on $[0, \frac{1}{2})$ and the angle between the lines of the grid and the tangent line to $\mathscr{C}$ at the given point is independently and uniformly distributed on $[0, 2\pi)$. Then the expected number of times that $\mathscr{C}$ will intersect the grid is given by $2L(\mathscr{C})/\pi$. A proof can be found in [**7**].

2. Experiments with concentric circles. We will now consider the INVERSE problem of designing an experiment that has a given distribution for the number of crossings. Formally, given a sequence $(p_0, p_1, p_2, \ldots)$ with $p_n \geqslant 0$ and $\Sigma_{n=0}^{\infty} p_n = 1$, we would like to design a curved needle experiment for which the probability of n crossings is p_n.

We begin with the case where the number of crossings must (with probability one) be an even number: i.e., $p_1 = p_3 = p_5 = \cdots = 0$. Our construction will be based on circles.

Jack Robertson has been at Washington State University since 1964 after receiving his Ph.D. in topology at the University of Utah under the direction of Eliot Chamberlin. His more recent research interests have been in the area broadly encompassed by combinatorial geometry. He is also active in mathematics education and enjoys rockhounding, beekeeping, gardening and backpacking.

Andy Siegel's interest in geometrical probability goes back to his Ph.D. research at Stanford University in 1977 under Herbert Solomon. His other interests include data analysis, robustness, graphics, Monte Carlo methods, time series analysis, computers, chi-squared distributions without any degrees of freedom, finance, zoology, hiking, bicycling, music, and family.

Let $\mathscr{C}$ denote the system of concentric circles $c_0, c_2, c_4, \ldots$ where the diameter of c_{2n} is

$$d_{2n} = 1 - \sum_{j=0}^{n} p_{2j},$$

and let x, $0 \leqslant x < \frac{1}{2}$, denote the distance from the common center of the circles to the nearest line of the grid. It is assumed that x is uniformly distributed on $[0, \frac{1}{2})$. In order that $2n$ crossings occur it is necessary and sufficient that $d_{2n}/2 < x < d_{2n-2}/2$ and this happens with probability $2 \cdot (d_{2n-2}/2 - d_{2n}/2) = p_{2n}$. Thus we have:

THEOREM 1. *Given a probability distribution on the even numbers* $0, 2, 4, \ldots,$ *there exists a system of concentric circles with this as the distribution of its number of crossings when thrown at random onto a system of parallel lines.*

3. The Buffon experiment with configurations of needles. We now return to the general problem except for a restriction on the probability of zero crossings. Assume a probability distribution

$$P = \{p_i\}_{i=0}^{\infty}, \quad p_i \geqslant 0, \quad \sum_{i=0}^{\infty} p_i = 1$$

is given with the restriction that $p_0 \geqslant 1 - 2/\pi$. Set

$$x_n = (\pi/2)\left(1 - \sum_{i=0}^{n} p_i\right), \quad n = 0, 1, 2, \ldots .$$

Then $x_n \leqslant 1$ for all n since $p_0 \geqslant 1 - 2/\pi$. This will insure the configuration we presently construct will intersect at most one line in the parallel grid which covers the plane. For each $n = 0, 1, 2, \ldots$ let L_n be a needle covering the interval $[0, x_n]$ (actually all that is required is that the needles be nested; they could as well share a common center). A configuration of stacked needles results with the interval $(x_n, x_{n-1}]$ covered by exactly n of the needles in our stack. Let this configuration of all the needles be kept together while cast randomly over a grid of horizontal parallel lines one unit apart. Let q be the common endpoint of all the needles of the cast configuration, let y be the distance from q to the grid line just above q, and let θ be the angle between that grid line and the ray from q which contains the configuration. We assume y is uniformly distributed on $[0,1)$ and θ is uniformly and independently distributed on $[0, \pi/2)$. By symmetry the cases $\theta \in [\pi/2, \pi)$, $[\pi, 3\pi/2)$ and $[3\pi/2, 2\pi)$ can be reduced to $\theta \in [0, \pi/2)$.

THEOREM 2. *When the needle configuration is cast on grid of parallel lines the probability of n crossings is* p_n.

Proof. Zero crossings occur when $x_0 \sin\theta < y$ and this happens with probability

$$(2/\pi)\int_0^{\pi/2}(1 - x_0 \sin\theta)\, d\theta = p_0 \quad \text{(Fig. 1)}.$$

Also n crossings occur when $x_n \sin\theta < y < x_{n-1}\sin\theta$ (described by the shaded region in Fig. 1) which happens with probability

$$(2/\pi)\int_0^{\pi/2}(x_{n-1} - x_n)\sin\theta\, d\theta = (2/\pi)(x_{n-1} - x_n) = p_n.$$

We finally look at the general case where p_0 might not exceed $1 - 2/\pi$. First a simple conditioned experiment can be used. For this purpose set $q_n = p_{n-1}$, $n = 1, 2, \ldots$. In this case let $x_0 = 1$ and

$$x_n = 1 - \sum_{i=1}^{n} q_i, \qquad n = 1, 2, \ldots .$$

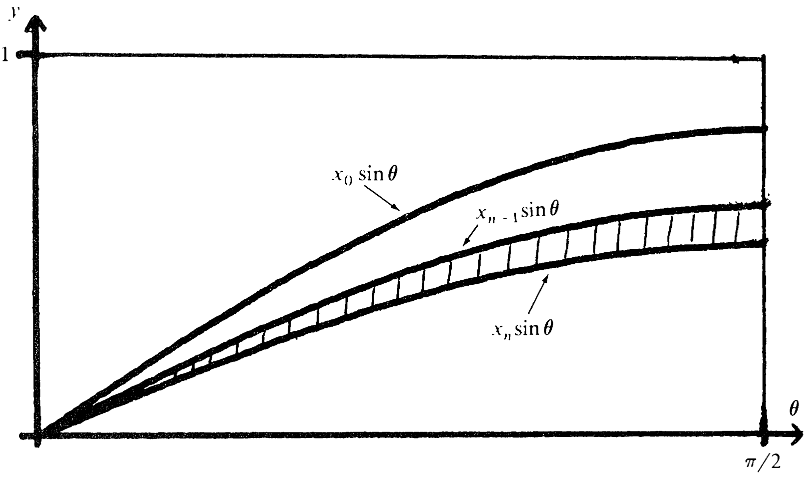


FIG. 1

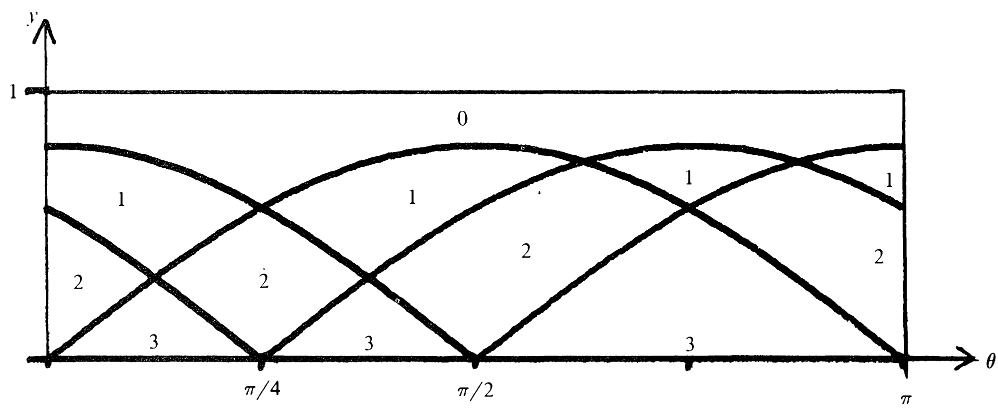


FIG. 2

For each n let L_n be a needle covering $[0, x_n]$ so that again n needles cover the interval $(x_n, x_{n-1}]$. Toss the needles and consider zero crossings a nonevent, i.e., let t_n be the conditional probability of n crossings, conditional upon there being at least one crossing, where $n = 1, 2, \dots$. As above it is easy to check that $t_n = q_n = p_{n-1}$ so that this experiment generates any given distribution.

Another approach is to use more general configurations in ways that the following example illustrates. Suppose there are three needles of length l, $0 < l < 1$, sharing an endpoint with angles such that two needles form a right angle while the third is at 45 degrees to these two. Conditions on y and θ are as above except now we let θ be uniformly distributed on $[0, \pi)$. The regions generating zero, one, two or three crossings are shown in Fig. 2. The procedure is easily adjusted to illustrate the generated distribution in cases where unequal angles lie between needles and the lengths of the needles vary and are allowed to exceed one in length. The generality realized is gained at the expense of more complicated computations of areas associated with the various probabilities.

References

1. M. E. Barbier, Note sur le problème de l'aiguille et jeu du joint couvert, J. Mathém., 5 (1860) 272–286.
2. G. Buffon, L'Histoire Naturelle, 4 Supplement (1777) 685–713.

3. D. Detemple and J. Robertson, Constructing Buffon curves from their distributions, this MONTHLY, 87 (1980) 779–784.
4. P. Diaconis, Buffon's problem with a long needle, J. Appl. Probab., 13 (1976) 614–618.
5. M. G. Kendall and P. A. P. Moran, Geometric Probability, Hafner, New York, 1963.
6. M. Perlman and M. Wichura, On sharpening Buffon's needle, Amer. Statist., 28 (1975) 157–163.
7. J. F. Ramaley, Buffon's noodle problem, this MONTHLY, 76 (1969) 916–918.
8. J. Roger, Georges-Louis Leclerc, Comte de Buffon, Dictionary of Scientific Biography, ed. C. Gillispie, Charles Scribner's Sons, New York, 1970.
9. E. F. Shuster, Buffon's needle experiment, this MONTHLY, 81 (1974) 26–29.
10. H. Solomon, Geometric Probability, SIAM, 1978.

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

A GROUP OF TWO PROBLEMS IN GROUPS

Rodney Forcade, Jack Lamoreaux & Andrew Pollington, Brigham Young University, Provo, UT 84602 observe that it is well known that the set $\{0,1,2,\dots,n\}$ can be made into an additive group by redefining ordinary addition: whenever $i+j>n$, replace $i+j$ by its least residue modulo $n+1$. They ask

Is it always possible to fix up multiplication in a similar way?

Is it possible, changing only those products which exceed n, to make the set $\{1,2,\dots,n\}$ into a multiplicative group? They conjecture that

¿ for every positive integer n there is a commutative group $(G,*)$ of order n, and a correspondence f from $\{1,2,\dots,n\}$ onto G such that whenever $ij \leqslant n$, then $f(ij)=f(i)*f(j)$. ?

Obviously, if $n+1$ is a prime, one can use multiplication, modulo $n+1$. If $2n+1$ is prime, let G be the group (of order n) of quadratic residues, modulo $2n+1$, and let f: $i \to i^2 \pmod{2n+1}$. Unfortunately, when $p=kn+1$ is prime, $k>2$, the map $i \to i^k \pmod p$ is not usually one-to-one. Perhaps, for every n, there is some prime $p=kn+1$, such that the kth powers of $1,2,\dots,n$ are distinct, modulo p?

The first n for which neither $n+1$ nor $2n+1$ is prime is $n=7$, but the conjecture is easily satisfied in this case. Let G be the cyclic group of order 7, generated by g, and put

$$f(2)=g,\quad f(3)=g^3,\quad f(5)=g^5,\quad f(7)=g^6,$$

so that

$$f(4)=g^2,\quad f(6)=g^4 \quad\text{and}\quad f(1)=g^7=1.$$

The authors have verified their conjecture for $n \leqslant 150$, using a computer. In every case it was possible to let G be a cyclic group and to let $f(2)$ be a generator. Perhaps that is always possible? Their conjecture is particularly interesting, because it implies

A conjecture of Graham.

¿ Given any set of more than $n \geqslant 1$ positive integers, there are always two elements a, b in the set with $a/\gcd(a, b) > n$. ?

Graham's conjecture has inspired many papers [**1**], [3], [7], [**9**], [**10**], [**11**], [**12**] and a thesis [**13**], not only because he offers \$25.00 for a solution. To see that the authors' conjecture implies Graham's, extend f naturally to a homomorphism from the positive integers onto G. By the pigeon-hole principle, if a set has more than n positive integers, then two of them have the same f-image. But if $f(a) = f(b)$, then $f(a/\gcd(a, b)) = f(b/\gcd(a, b))$ which implies that either $a/\gcd(a, b) > n$ or $b/\gcd(a, b) > n$, since f is one-to-one on $\{1, 2, \ldots, n\}$.

The authors include the word "commutative" in their conjecture because, without it, the implication of Graham's conjecture fails, and because they believe, though they haven't proved, that every group satisfying their conjecture must be commutative. For example, when $n = 12$, it is possible to satisfy the conjecture, using either of the commutative groups of order 12, but it is easy to see that no non-commutative group can work for $n = 12$: first, $f(2)$ commutes with $f(1), f(2), \ldots, f(6)$, because $2 \cdot 1, 2 \cdot 2, \ldots, 2 \cdot 6 \leq 12$. Also $f(2) \cdot f(8) = f(2) \cdot f(4) \cdot f(2) = f(8) \cdot f(2)$, so $f(2)$ commutes with more than half of the elements, and hence with the entire group. A similar argument can be made for $f(3)$, so the centre contains all elements generated by 2 and 3.

[**Added in proof**: Mário Szegedy [**14**] has produced a complicated proof of Graham's conjecture, which Szemerédi and others have examined. It will probably be published in the Hungarian periodical *Combinatorica*.]

Richard D. Feuer, 230 West End Avenue, New York, NY 10023 poses

A problem in permutation group theory.

Let F be the free group on two generators α, β, and G be the symmetric group of $\{1, 2, \ldots, n\}$. Given elements a, b of G, there is a unique homomorphism, $F \to G$, which takes α to a and β to b. Let $f(a, b)$ denote the image of the element f of F under this homomorphism. For example, if $f = \alpha^2\beta\alpha^{-1}$, then $f(a, b) = a^2ba^{-1}$.

With the exception of the case $n = 6$ [**2**], [**8**], every automorphism of G is an inner automorphism, i.e., is conjugation by some element of G. Let a, b, A, B be elements of G. If there is an inner automorphism of G carrying A to a and B to b, then, plainly, it must carry each $f(A, B)$, with f in F, to the corresponding element $f(a, b)$ and so

$$(*) \qquad f(A, B) \text{ and } f(a, b) \text{ are conjugate in } G, \qquad \text{for all } f \text{ in } F.$$

Conversely, if A, B, a, b satisfy condition $(*)$, then it is easy to verify that there is an isomorphism σ of the subgroup generated by A, B to that generated by a, b which carries A to a and B to b. Feuer conjectures that σ can always be extended to an inner automorphism of G. He has verified this conjecture when the number of integers moved by A, B is at most one.

References

1. R. D. Boyle, On a problem of R. L. Graham, Acta Arith., 34 (1978) 163–177.

2. W. Burnside, Theory of Groups of Finite Order, 2nd ed., Cambridge University Press, 1911; reprinted, Dover, 1955; Chapter XI, §162.

3. E. Z. Chein, On a conjecture of Graham concerning a sequence of integers, Canad. Math. Bull., 21 (1978) 285–289; MR 80d:10024; Zbl 392.10002.

4. P. Erdős and R. L. Graham, Old and New Problems and Results in Combinatorial Number Theory, Monographies de l'Enseignement Math., No. 28, Geneva, 1980, p. 78.

5. R. L. Graham, Advanced problem 5749*, this MONTHLY, 77 (1970) 775.

6. Richard K. Guy, Unsolved Problems in Number Theory, Springer-Verlag, New York, 1981, Problem F15.

7. J. Marica and J. Schönheim, Differences of sets and a problem of Graham, Canad. Math. Bull., 12 (1969) 635–637.

8. D. W. Miller, On a theorem of Hölder, this MONTHLY, 65 (1958) 252–254.

9. R. J. Simpson, On a conjecture of R. L. Graham, Acta Arith., 40 (1982) 209–211.

10. William Yslas Vélez, Some remarks on a number theoretic problem of Graham, Acta Arith., 32 (1977) 233–238.

11. Gerald Weinstein, On a conjecture of Graham concerning greatest common divisors, Proc. Amer. Math. Soc., 63 (1977) 33–38; Zbl. 369.10003.

12. Riko Winterle, A problem of R. L. Graham in combinatorial number theory, Congressus Numerantium, I, Proc. Conf. Combin. Baton Rouge, 1970, 357–361; MR 42#305.

13. W. W. C. Wong, On a number theoretic conjecture of Graham, Ph.D. thesis, Columbia Univ., 1981.

14. Márió Szegedy, The solution of Graham's common divisor problem, Preprint No. 27/1985, Math. Inst. Hungar. Acad. Sci., Reáltanoda u. 13–15, Budapest H-1053.

NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.

For instructions about submitting Notes for publication in this department see the inside front cover.

TWO ELEMENTARY GENERALISATIONS OF BOOLEAN RINGS

MÍCHEÁL Ó SEARCÓID AND DESMOND MACHALE

Department of Mathematics, University College, Cork, Ireland

It is well known that a Boolean ring, i.e., a ring in which every element x satisfies $x^2 = x$, is commutative. In this note we generalise this result by showing that a ring R is commutative if it satisfies either of the following conditions:

1. All products in R are idempotent, i.e. $(xy)^2 = xy$, for all $x, y \in R$.
2. All commutators in R are idempotent, i.e. $(xy - yx)^2 = xy - yx$, for all $x, y \in R$.

Condition 1 is certainly weaker than the condition $x^2 = x$, for all $x \in R$, since there exist non-Boolean rings satisfying $(xy)^2 = xy$, for all x and y. As an example, consider any ring with trivial product, $xy = 0$ for all x and y. Condition 2 is obviously also necessary for commutativity and is in fact a special case of a celebrated theorem of Herstein [**1**]. However, the proof we present is completely elementary.

Notation and terminology are standard. R is an associative ring; an idempotent is an element e in R with $e^2 = e$, and $Z(R)$, the centre of R, is the set $\{r \in R | rx = xr, \text{ for all } x \in R\}$. We need the following lemma.

LEMMA. *Let R be a ring in which $yx = 0$ implies $xy = 0$, for $x, y \in R$. If e is an idempotent in R, then $e \in Z(R)$.*

Proof. For all $r \in R$, $(e^2 - e)r = 0 = e(er - r)$. By hypothesis, $(er - r)e = 0$ so $ere = re$. Similarly, $(re - r)e = 0$ implies $e(re - r) = 0$, giving $ere = er$. Thus $er = re$ and so $e \in Z(R)$.

THEOREM 1. *If for all $x, y \in R, (xy)^2 = xy$, then R is commutative.*

Proof. If $yx = 0$, then $xy = (xy)^2 = x(yx)y = 0$, so, by the lemma, $xy \in Z(R)$ for all

$x, y \in R$. Thus, since all products are central,

$$xy = (xy)^2 = x(yxy) = (yxy)x = (yx)^2 = yx,$$

and so R is commutative.

THEOREM 2. *If for all* $x, y \in R, (xy - yx)^2 = xy - yx$, *then* R *is commutative.*

Proof. If $yx = 0$, then $xy = (xy)^2 = x(yx)y = 0$, so, by the lemma, $xy - yx \in Z(R)$, for all x and y. Further,

$$xy - yx = (xy - yx)^2 = (yx - xy)^2 = yx - xy,$$

so $x(xy - yx) = (yx - xy)x$. This gives at once that $x^2 \in Z(R)$, for all $x \in R$. Next,

$$\begin{aligned}(xy)^2 &= x(yxy) = x\left[(yx)^2 + y^2 - (yx - y)^2 - y^2x\right] \\ &= \left[(yx)^2 + y^2 - (yx - y)^2 - y^2x\right]x = (yxy)x = (yx)^2,\end{aligned}$$

for all $x, y \in R$. Finally,

$$\begin{aligned}xy - yx &= (xy - yx)^2 = xy(xy - yx) - yx(yx - xy) \\ &= (xy)^2 - xy^2x - (yx)^2 + yx^2y = -x^2y^2 + x^2y^2 = 0,\end{aligned}$$

which means that R is commutative.

REMARKS. One might conjecture the following possible generalizations of Theorem 1 and 2.

(i) Let R be a ring in which $(xyz)^2 = xyz$ for all $x, y, z \in R$. Then R is commutative.
(ii) Let R be a ring in which $(xy)^2 - xy \in Z(R)$ for all $x, y \in R$. Then R is commutative.
(iii) Let R be a ring in which $(xy - yx)^2 - (xy - yx) \in Z(R)$ for all $x, y \in R$. Then R is commutative.

However, the non-commutative ring M of strictly upper triangular 3×3 matrices over Z_3 satisfies the condition $xyz = 0$ for all $x, y, z \in R$ and so is a counterexample to all three conjectures.

In conclusion, we note that Theorem 1 is a special case of the following more general result:

If for each $x, y \in R$, *there exists* $n = n(x, y) > 1$ *such that* $(xy)^n = xy$, *then* R *is commutative.*

Since we are not aware of any readily accessible source of this result in the literature, we indicate here the outline of a proof, which, however, uses relatively heavy machinery.

In a ring R satisfying the given condition, clearly $xy = 0 \Rightarrow yx = 0$. It follows that if x is nilpotent and y is arbitrary, then xy nilpotent. Thus, the nilpotent elements of R annihilate R on both sides, and are therefore central. Since R is obviously periodic, commutativity now follows by a result of Bell [**2**].

References

1. I. N. Herstein, Non-Commutative Rings, Carus Mathematical Monograph No. 15, Mathematical Association of America, Washington, DC, 1968.
2. H. E. Bell, A Commutativity Study for Periodic Rings, Pacific J. Math., 70 (1977) 29–36.

ANSWER TO PHOTO ON PAGE 105

Constantin Carathéodory, 1873–1950, author of many papers on geometric optics and of the immortal *Vorlesungen über reelle Funktionen*.

Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

T:	Textbook	13-18:	Grade Level
S:	Supplementary Reading	1-4 :	Time in Semesters
P:	Professional Reading	** :	Special Emphasis
L:	Undergraduate Library	?? :	Questionable

General, P. Current Topics in Chinese Science, Section C: Mathematics, Volume 3 (1984). Gordon & Breach, 1984, 869 pp, $98 (P). [ISBN: 0-677-40405-0] A reprinting (in English) of over 150 selected mathematics papers that appeared in 1983 in Science in China (Scientia Sinica) and in Science Bulletin (Kexue Tongbao). The third such reprint volume produced in this series. (Similar series are published in six other sciences.) LAS

General, L. Shogi: Japan's Game of Strategy. Trevor Leggett. Charles E Tuttle, 1982, 100 pp, $12.95. [ISBN: 0-8048-0526-1] Shogi is the Japanese representative in the family of chess games. Eighth printing, first published in 1966. Cardboard Shogi board and punch-out Shogi pieces are included. Step-by-step instructions with numerous helpful diagrams. For seasoned players as well as beginners. JK

General, S. ALEX. Ian D. Macdonald. Polygonal, 1984, (P). [ISBN: 0-936428-10-4] A sophomoric fantasy of a young Euclid's decision to move from Athens to Alexandria, complete with imagined but pointless dialogue about the impractical dreamings of the youthful mathematician. LAS

General, S(13-16), L. Dictionary of Mathematics. Ed: J.A. Glenn, G.H. Littler. Barnes & Noble, 1984, x + 230 pp. [ISBN: 0-389-20451-X] A compilation of terms from indexes of British undergraduate textbooks, closed under definition so that each technical word used in the dictionary is defined therein. Emphasizes more geometry (especially algebraic curves) and less analysis than is common in U.S. curricula. Does not cover computer science. LAS

General, S(16-17), P. Formulaire de mathématiques. L. Chambadal. Dunod, 1985, 188 pp, (P). [ISBN: 2-04-015759-X] The mathematics (definitions, theorems, formulas) covered in preparation for scientific education at French universities. (First Edition, TR, June-July 1973; Second Edition, TR, January 1979.) JD-B

General. The Technical Editor's and Secretary's Desk Guide. George Freedman, Deborah A. Freedman. McGraw-Hill, 1985, xxxviii + 540 pp, $29.95. [ISBN: 0-07-021918-4] A useful handbook for anyone preparing manuscripts using college-level mathematics and science. Includes a bit of substance about each symbol to help the non-mathematical practitioner understand how it is used. Covers basic mathematics, physics, electronics, and chemistry. LAS

General, S(13-14), L. The Garden of the Sphinx: 150 Challenging and Instructive Puzzles. Pierre Berloquin. Transl: Charles Scribner, Jr. Charles Scribner, 1985, xvii + 185 pp, $13.95. [ISBN: 0-684-18342-0] 150 clever puzzles from a column in Le Monde, requiring only school mathematics for their solution. Sample: Fold a square piece of paper to form a regular hexagon. Complete solutions are provided in the second half. LAS

General, S(13-14), L. Problèmes résolus de mathématiques. Saint-Martin. Dunod, 1985, 250 pp, (P). [ISBN: 204-015780-8] Problems and solutions from 55 regional exams given in France to candidates of "General Mathematics" A_1 and A_2. Five to seven multipart questions per each 3 to 4 hour exam. Interesting problems on calculus (single and multivariable), linear algebra, elementary differential equations and complex variables. RM

General, S, L*. Chinese Chess. H.T. Lau. Charles E Tuttle, 1985, 248 pp, $11.50. [ISBN: 0-8048-1495-3] Chinese Chess, also known as "elephant" chess, is a variant of the Western game that emerged about 1000 years ago when both Western and Eastern versions branched from a common origin in Persia. This book explains the pieces and their moves, then discusses standard openings, mid-game strategy, and numerous end games. Includes in an appendix many complete games recorded in the two earliest known Chinese manuscripts on chess, from the Ming Dynasty. LAS

Elementary, T*(13), S. Algebra for College Students. R. David Gustafson, Peter D. Frisk. Brooks/Cole, 1985, xiii + 546 pp, $30.25. [ISBN: 0-534-05028-X] Essentially, an intermediate algebra

course which includes a few topics that are more advanced. Includes lots of worked examples, many exercises, and a thorough review of basic concepts. Color is used to highlight important ideas. Written with the student in mind. CEC

Elementary, T(13: 1). Intermediate Algebra. Michael N. Payne. West, 1985, xiii + 577 pp, $27.95. [ISBN: 0-314-85285-9] Traditional range of material with abundant exercises (to an extreme). Importance of problem solving is highlighted although the majority of problems follow traditional mold (age, mixture, rate/time). Good use of graphics to support algebra concepts; text includes highlighted procedures and cautions throughout, with concise chapter reviews. RD

Elementary, T(13: 1). Elementary Algebra for Today. Donald E. Brook. Prentice-Hall, 1985, x + 404 pp, $23.95. [ISBN: 0-13-252842-8-01] Topics proceed from signed numbers to quadratic equations; each new topic is motivated with true-to-life practical applications. Incorporates reviews of arithmetic concepts and stresses visual images of number relationships. Attractive, open format; conversational style with "reading questions" throughout the textual material. LCL

Elementary, T(13: 1). Data Processing Mathematics. Elizabeth Bliss. Prentice-Hall, 1985, x + 254 pp, $24.95. [ISBN: 0-13-196155-1] Presupposes elementary algebra. Number bases, set language, elementary logic, flowcharts, matrix arithmetic, sequences, linear programming, and logarithms. The text is intended to "provide data processing students with...mathematical topics relevant to the field of computers." FLW

Mathematics Appreciation, T(13-16: 2), S, L. Mathematics for the Nonmathematician. Morris Kline. Dover, 1985, xiii + 641 pp, $10.95 (P). [ISBN: 0-486-24823-2] Unabridged republication of Mathematics for Liberal Arts (Addison-Wesley, 1967), including, at the end, the original typewritten Instructor´s Manual. Deals predominantly with the history of mathematics and physical sciences, covering modern topics and other applications only briefly near the end. An excellent resource, but by now somewhat dated in tone and selection of topics. LAS

Mathematics Appreciation, S(13-16), L*. Wheels, Life and Other Mathematical Amusements. Martin Gardner. WH Freeman, 1983, ix + 261 pp, $9.95 (P). [ISBN: 0-7167-1589-9] Paperback reprinting of the 1983 hardcover edition containing updated reprints of 22 columns from Scientific American. Includes all three Gardner columns on Conway´s game of "Life," as well as chapters on other mathematical games (Nim, Halma, Hackenbush, etc.). LAS

Mathematics Appreciation, T(13-16: 1), S, L*. Mathematics in Civilization. H.L. Resnikoff, R.O. Wells, Jr. Dover, 1984, viii + 408 pp, $9.95 (P). [ISBN: 0-486-24674-4] Reprint of the 1973 original edition (TR, March 1974), supplemented with a brief final chapter on twentieth century mathematics and an appendix with solutions to the exercises. Provides a substantial, historically-organized sketch of the role played by mathematics in the rise of measurement, from prehistoric counting to contemporary models of the universe. LAS

Mathematics Appreciation, S(13), L. 3.1416 And All That. Philip J. Davis, William G. Chinn. Birkhauser Boston, 1985, ix + 188 pp, $14.95 (P). [ISBN: 0-8176-3304-9] 24 brief, self-contained vignettes of elementary mathematics, reprinted from columns by the authors in Science World, a news magazine for school students and teachers. Reprint of the 1969 Simon & Schuster First Edition (TR, November 1970) supplemented with a brief update on recent calculations of pi. LAS

Mathematics Appreciation, S(13-18), P, L. The Magic Numbers of Dr. Matrix. Martin Gardner. Prometheus Books, 1985, 326 pp, $19.95; $10.95 (P). [ISBN: 0-87975-281-5; 0-87975-282-3] Dr. Matrix said "Numbers, you know, have a mysterious life of their own." No one reveals their secret life as well as does Martin Gardner. All 22 Dr. Matrix columns are collected here. The first 18 appeared in The Incredible Dr. Matrix, published in 1976. In Chapter 22, Gardner reports the death of Dr. Matrix, "the strangest and the wisest man I have ever known." JK

Precalculus, T(13: 1). College Algebra with Applications. José Barros-Neto. West, 1985, xi + 526 pp [ISBN: 0-314-85217-4]; College Algebra and Trigonometry with Applications, xiii + 704 pp. [ISBN: 0-314-85218-2] Standard college text. Few special applications; little to be enthusiastic about. The second text has three trigonometry chapters, otherwise the two are identical. RD

Precalculus, T(13: 1). College Algebra: A Straightforward Approach, Second Edition. Martin M. Zuckerman. Wiley, 1985, xvii + 515 pp, $25.95. [ISBN: 0-471-09619-9] Offers many examples and abundant real-world applications. Graphs, charts, and drawings enliven text. Systematic progression with emphasis on learning skills. RD

Precalculus, T(13: 1). College Algebra, Second Edition. John R. Durbin. Wiley, 1985, ix + 510 pp, $25.95. [ISBN: 0-471-81714-7] Nothing more than a standard introductory text; mundane applications including endless number of compound interest problems. Problem solving never mentioned nor attempted. First four chapters of earlier edition (TR, October 1982) compressed into two, otherwise little change. RD

Precalculus, T*(13: 1). Precalculus. Roland E. Larson, Robert P. Hostetler. DC Heath, 1985, xvi + 597 pp, $26.95. [ISBN: 0-669-08617-7] Attractive presentation of algebraic and trigonometric functions and analytic geometry. Emphasizes a graphical approach and algebraic problems from calculus. Includes techniques and problems for calculator use. JNC

Finite Mathematics, T(13), S, L. Mathematics for Data Processing and Computing. Maria Shopay Kolatis. Addison-Wesley, 1985, xviii + 549 pp, $25.95. [ISBN: 0-201-14955-9] A mathematics text designed specifically for the data processing or computer science student. Part 1 develops problem-solving techniques and topics such as sets, equations and permutations. Part 2 covers symbolic logic and Boolean algebra from a computer scientists point of view. Part 3 deals with computer number systems, and Part 4 looks at sequences, matrices and computer error. Readable with good problem sets. CEC

Education, P*, L. Second International Mathematics Study: Summary Report for the United States. Kenneth J. Travers. Stipes, 1985, xix + 97 pp, (P). [ISBN: 0-87563-267-X] Revision of a preliminary report issued in September 1984, documenting a 1981-82 international study that placed U.S. twelfth grade students in the lowest quartile internationally. LAS

History, S(15-17), P*, L. Nicolas Chuquet, Renaissance Mathematician. Ed: Graham Flegg, Cynthia Hay, Barbara Moss. D Reidel, 1985, viii + 388 pp, $49. [ISBN: 90-277-1872-5] Critical assessment of a large portion of Chuquet's mathematical manuscript (325 folios) completed in 1484, including not only his best-known work, the Triparty (in brief, arithmetic, roots, and algebra), but also his Geometry and Commercial Arithmetic. Chapter 9 is a pass at assessing the importance of Chuquet's writings relative to those of Fibonacci and Pacioli. Appendix contains the table of contents for Chuquet's mathematical manuscripts. This scholarly work will be of special interest to historians of mathematics and Renaissance historians. JK

History, T(15-17), S, L. A History of Mathematics. Carl B. Boyer. Princeton U Pr, 1985, xv + 717 pp, $12.50 (P). [ISBN: 0-691-02391-3] First paperback edition of the 1968 classic text, originally published by John Wiley (TR, January 1969). LAS

History, P*. The Astronomy of Levi ben Gerson (1288-1344). Bernard R. Goldstein. Springer-Verlag, 1985, xi + 308 pp, $68. [ISBN: 0-387-96132-1] Translation of and commentary on Chapters 1-20 of The Astronomy of Levi Ben Gerson (1288-1344). Except for the table of contents and some poems celebrating the invention of the Jacob Staff, this is the first edition of the Hebrew text. Some pieces of the commentary have appeared earlier in papers by Goldstein. Scholarly work with numerous references. Brief section on Levi and his predecessors and on his influence on subsequent astronomers. JK

History, P*, L*. Alfred North Whitehead: The Man and His Work, Volume I: 1861-1910. Victor Lowe. Johns Hopkins U Pr, 1985, xi + 351 pp, $27.50. [ISBN: 0-8018-2488-5] First of two volumes of the first full biography of Whitehead, this covering his years as mathematics lecturer at Trinity College. Scant written sources from his early years--Whitehead was an intensely private person--make this well-researched volume especially welcome. Covers Whitehead's collaboration with Bertrand Russell, both in professional and personal detail. LAS

Foundations, P. Penser les mathématiques. R. Apéry, et al. Editions du Seuil, 1982, 273 pp, (P). [ISBN: 2-02-006061-2] Twelve papers from the seminar on philosophy and mathematics of l'Ecole normale supérieure, with contributions from Apéry, Dieudonné, Loi, Mandelbrot, Thom, et al. Divided into two thematic groups--mathematics and language, mathematics and nature. Extensive editors' notes supplement individual bibliographies to connect these essays with related literature. LAS

Graph Theory, S(17), P, L. Convexity and Graph Theory. Ed: M. Rosenfeld, J. Zaks. Math. Stud., V. 87. Elsevier Sci, 1984, xi + 339 pp, $59 (P). [ISBN: 0-444-86571-3] Proceedings of a conference held in Israel in March, 1981. Includes papers given by some of the best-known graph theorists and geometers. CEC

Combinatorics, S(15-18), L**.** Handbook of Applicable Mathematics, V. V: Combinatorics and Geometry. Ed: Walter Ledermann, Steven Vajda. Wiley, 1985, $170 set [ISBN: 0-471-90023-0]; Part A, xxxiii + 327 pp; Part B, xxxiii + 403 pp. In two parts, the fifth of six volumes whose aim is "to provide up-to-date and easily accessible information about the practical aspects of all mathematics." Other volumes cover algebra, probability, numerical methods, analysis and statistics. Topics have been chosen for their relevance to a number of different fields of application. Great for browsing; excellent for whetting the appetites of curious students. Volume 5 covers trigonometry (plane and spherical), curve sketching, topology, convexity, graphs and networks, tensors, catastrophe theory, finite spaces and combinatorial designs, projective geometry, symmetry, differential geometry, analytic manifolds and Lie groups, enumeration, coding theory and some of the latest results obtained by Grünbaum and Shepherd on patterns. The list of authors is an all-star array. JK

Discrete Mathematics, P. Advances in Discrete Mathematics and Computer Science, Volume I: Neofields and Combinatorial Designs. Ed: D. Frank Hsu. Hadronic Pr, 1985, xii + 396 pp, $60 (P). [ISBN: 0-911767-26-6] The first of a series that plans to collect important papers for research workers in various special areas of discrete mathematics and computer science. The 29 papers in this volume begin with L.J. Paige's 1949 work on neofields and then continues with papers on developments and extensions that deal with all sorts of algebraic, combinatorial and geometric structures. SS

Number Theory, P. Lecture Notes in Mathematics-1122: Number Theory. Ed: K. Alladi. Springer-Verlag, 1985, vii + 217 pp, $14.40 (P). [ISBN: 0-387-15222-9] Conference held in Ootacamund, India, January 1984, coinciding with Paul Erdos' 70th birthday. Contains 17 papers, including an address by Erdos titled "On some of my problems in number theory I would most like to see solved." BC

Number Theory, P. Lecture Notes in Mathematics-1135: Number Theory. Ed: D.V. Chudnovsky, et al. Springer-Verlag, 1985, 283 pp, $17.60 (P). [ISBN: 0-387-15649-6] Second volume of proceedings from the New York Number Theory Seminar (Volume I = SLN 1052). Non-number theorists take note: these proceedings include papers on Hausdorff dimension, Pade approximation, partial differential equations, K3 surfaces, trigonometric polynomials, and the Brauer group. BC

Number Theory, S*(11-14), L. Adventures with Your Computer. L. Råde, R.D. Nelson. Penguin Books, 1984, 202 pp, (P). [ISBN: 0-14-007362-0] Sixteen groups of number exercises (combinatorics, patterns, random numbers) intended for the reader with a simple computer. The middle section--the bulk of the book--consists of commentary on the exercises, suggesting lines of attack, connections to other ideas, and references. An ideal source of inspiration for discovery teaching. LAS

Linear Algebra, T(14-15: 1, 2), L. Elementary Linear Algebra, Second Edition. A. Wayne Roberts. Benjamin/Cummings, 1985, xv + 414 pp. [ISBN: 0-8053-8305-0] Second Edition covers standard topics of linear algebra plus a chapter each on applications to linear programming, Markov chains, and differential equations. (Eigenvalues though are given ten pages, with the emphasis on diagonalization of the matrix.) Includes a mix of computational and "theoretic" problems. Best feature: "vignettes" showing interesting uses of linear algebra. BC

Group Theory, P. Lecture Notes in Mathematics-1112: Products of Conjugacy Classes in Groups. Ed: Z. Arad, M. Herzog. Springer-Verlag, 1985, 244 pp, $14.40 (P). [ISBN: 0-387-13916-8] Proves a covering theorem (with extensions) for a group G guaranteeing the existence of an n such that $C^n = G$ (respectively, $C_1 \cdot C_2 \cdots C_n = G$) for any conjugacy class $C \neq e$ (respectively, $C_i \neq e$). Four independent chapters give recent results on such theorems for important classes of simple groups. RM

Group Theory, P. Abelian Groups and Modules. Ed: R. Göbel, et al. Springer-Verlag, 1984, xii + 531 pp, $40.60 (P). [ISBN: 0-387-81847-2] Proceedings of an April 1984 conference in Udine, Italy including most of the 11 invited lectures but excluding (for reasons of space) all survey papers. LAS

Algebra, T(18: 1), S, P. Orthomodular Lattices: Algebraic Approach. Ladislav Beran. Math. & Its Applic. D Reidel, 1985, xix + 394 pp, $69. [ISBN: 90-277-1715-X] Intended as an introductory text at the graduate level; assumes some background in the theory of ordered sets. Introductory chapters on orthomodular lattices are followed by chapters on amalgams, generalized orthomodular lattices, solvability and special properties, and applications, including dimension theory. Exercises (with solutions), references, index. JS

Algebra, P. Representation of Rings Over Skew Fields. A.H. Schofield. London Math. Soc. Lecture Note Ser., V. 92. Cambridge U Pr, 1985, xii + 223 pp, $27.95 (P). [ISBN: 0-521-27853-8] A classification of all finite dimensional representations of rings over skew fields, followed by a detailed study of specific skew fields, especially those arising in the work of P.M. Cohn. LAS

Algebra, S(17), P. Modular Forms. Ed: Robert A. Rankin. Math. & Its Applic. Halsted Pr, 1984, 272 pp, $54.95. [ISBN: 0-470-20099-5] This book assembles the contributions to a symposium on modular forms of one and several variables held at the University of Durham, June 30 to July 10, 1983. Includes accounts of some of the important new results which were obtained in the early part of this decade. CEC

Algebra, T(18: 1), S, P. Projective Representations of Finite Groups. Gregory Karpilovsky. Pure & Appl. Math., V. 94. Dekker, 1985, xiii + 644 pp, $89.75. [ISBN: 0-8247-7313-6] A self-contained selective account of projective representations based on Schur's 1904-1911 work but emphasizing results based on the study of modules over twisted group algebras. Prerequisite: a solid year graduate course in algebra. LAS

Algebra, S(18), P. Growth of Algebras and Gelfand-Kirillov Dimension. G.R. Krause, T.H. Lenagan. Research Notes in Math., V. 116. Pitman, 1985, 182 pp, $16.95 (P). [ISBN: 0-273-08662-6] These notes aim to "present in an orderly fashion as much as possible of the basic material concerning Gelfand-Kirillov dimension" of an algebra together with several significant applications. A general introduction is followed by later applications to Weyl algebras, enveloping algebras for Lie algebras, polynomial identity algebras, and growth of groups. References, index. JS

Algebra, T(17-18: 1, 2), P. Linear Groups and Permutations. A.R. Camina, E.A. Whelan. Research Notes in Math., V. 118. Pitman, 1985, 152 pp, $15.95 (P). [ISBN: 0-273-08672-3] Self-contained study of the structure, representations, and classification of finite simple linear groups. Ends with exposition of C. Hering's results, with applications to permutation groups and automorphism groups of finite geometries and block designs. RM

Algebra, S, P. Modules Over Valuation Domains. Laszlo Fuchs, Luigi Salce. Lect. Notes in Pure & Appl. Math., V. 97. Dekker, 1985, xi + 317 pp, $55 (P). [ISBN: 0-8247-7326-8] Modules over valuation domains represent a middle road between the well-traveled theory of abelian groups and the uncharted theory of modules over arbitrary domains. This volume is the result of the authors' search for a systematic treatment of modules over valuation domains. Many of the ideas and methods are inspired by the authors' considerable research and experience with abelian group theory. LCL

Calculus, T(13: 2-3), S**, L.** Methods of Mathematics Applied to Calculus, Probability, and Statistics. Richard W. Hamming. Ser. in Comput. Math. Prentice-Hall, 1985, xviii + 857 pp, $49.95. [ISBN: 0-13-578899-4] A most unconventional calculus book: the author's "labor of love." Based on

"regeneration rather than retrieval of knowledge." Stresses understanding, abstraction, and methods of reasoning. Includes some probability, some statistics and more than usual emphasis on discrete mathematics. Excellent prologue and epilogue on the importance, uniqueness and effectiveness of mathematics, with a brief piece on rigor (which the author calls the hygiene of mathematics--what is needed to protect us against careless thinking). Every chapter contains tidbits which every calculus teacher will savor. This textbook might not fit most school's calculus sequences but will well serve those which can afford to offer a leisurely-paced one-year course in which one can dwell on those topics which strike one's fancy. JK

Calculus, S. Mathematics for Chemists. P.G. Francis. Chapman & Hall, 1984, x + 193 pp, $15.95 (P); $33. [ISBN: 0-412-24990-1; 0-412-24980-4] A potpourri of topics from analytic geometry, calculus, vectors, differential equations and statistics. Intended to prepare (British) chemistry students for first-degree courses. Useful, perhaps, but note the price! TAV

Calculus, T(13-14: 1).** Calculus III, Second Edition. Jerrold Marsden, Alan Weinstein. Undergrad. Texts in Math. Springer-Verlag, 1985, xiv + 338 pp, $17.95 (P). [ISBN: 0-387-90985-0] Text for third or fourth-quarter calculus. Presumes no prior knowledge of linear algebra. Introduces partial derivatives, gradients, Lagrange multipliers, multiple integrals, and Stokes' theorem among other topics. Includes both theoretical discussions and many examples of applications. Nice illustrations. AM

Real Analysis, S(18), P. Classical Real Analysis. Ed: Daniel Waterman. Contemp. Math., V. 42. AMS, 1985, ix + 216 pp, $22 (P). [ISBN: 0-8218-5045-8] A collection of papers based on a special session on classical analysis in honor of Casper Goffman. A variety of topics, including Goffman's talk on Cesari and Sobolev spaces as related to multiple Fourier series. JS

Real Analysis, S(18), P. The Theory of Functions of a Real Variable. R.L. Jeffery. Dover, 1985, xiii + 232 pp, $6 (P). [ISBN: 0-486-64781-1] An unabridged and unaltered version, in paperback, of the Second Edition (1953). Largely a study of classical integration theory, particularly the Lebesgue theory. Bibliography, indexes. JS

Real Analysis, T(15-16: 1), L. Introduction to Real Variable Theory. Subhash Chandra Saxena, S.M. Shah. Prentice-Hall, 1980, xv + 334 pp, $19.95 (P). [ISBN: 0-13-494030-X] Some additional diagrams and exercises have been added in this revised edition (1972 Intext Publishers edition, TR, October 1972). Nicely written, with plenty of topics to choose from. Over a quarter of the book concerns series (including Fourier series). Riemann integration; measure theory and Lebesgue integration. Much point set theory. DA

Complex Analysis, S(18), P. Analytic Functions--Growth Aspects. O.P. Juneja, G.P. Kapoor. Research Notes in Math., V. 104. Pitman, 1985, 296 pp, $21.95 (P). [ISBN: 0-273-08630-8] A research monograph, extending the mature theory of growth behavior of entire functions to the case of analytic functions on finite domains. The treatment is principally in one, but also in several complex variables. Relatively self-contained; basic results are proved. Includes hundreds of exercises. PZ

Complex Analysis, T(15-16: 1), S, L. A First Course in Applied Complex Variables. Lester A. Rubenfeld. Wiley, 1985, x + 470 pp, $34.95. [ISBN: 0-471-09843-4] Concrete, intuitive, relatively elementary treatment, emphasizing applications: fluid flow, electrostatics, heat conduction, conformal mappings. Theoretical topics (integration theorems, series developments) appear relatively late. Exercises are mostly computational; nearly all are solved in the appendix. PZ

Complex Analysis, T(16-17: 1, 2). Complex Analysis, Second Edition. Serge Lang. Grad. Texts in Math., V. 103. Springer-Verlag, 1985, xiv + 367 pp, $39. [ISBN: 0-387-96085-6] Several sections have been rewritten and some new material added. (First Edition, TR, February 1979.) MU

Complex Analysis, S(18), P. Blaschke Products: Bounded Analytic Functions. Peter Colwell. U of Michigan Pr, 1985, viii + 140 pp, $15. [ISBN: 0-472-10065-3] A detailed overview, with commentary, of the theory of Blaschke products from its inception in 1915 to the present. Results are collected from over 300 papers in bibliography; no proofs are included. PZ

Complex Analysis, T(18: 1), S, P. Univalent Functions. A.W. Goodman. Mariner, 1983 [ISBN: 0-936166-12-6]. Volume I, xvii + 246 pp; Volume II, xii + 311 pp. A readable, leisurely introduction to current research and to more advanced monographs in the area. With hundreds of exercises. Aimed at graduate students and non-specialists--either volume could serve as one-term textbook. Omits Loewner theory and Grunsky inequalities (important in recent proof of the Bieberbach conjecture). Contains an historical afterword. PZ

Differential Equations, P. Points Fixes, Points Critiques et Problèmes aux Limites. Jean Mawhin. Pr U Montreal, 1985, 162 pp, $19 (P). [ISBN: 2-7606-0696-1] Carefully-prepared lecture notes on the analysis of second-order nonlinear (ordinary) differential equations with periodic solutions, such as forced pendulum equations. Appendices on Fredholm operators, convex analysis, and critical points. Good historical introduction. BC

Differential Equations, S(17-18). Bifurcation Analysis: Principles, Applications and Synthesis. Ed: M. Hazewinkel, R. Jurkovich, J.H.P. Paelinck. D Reidel, 1985, vii + 259 pp, $38. [ISBN: 90-277-1446-0] Consider a differential equation depending on a parameter. Bifurcation theory describes how the qualitative behavior of solutions changes as the parameter varies. This book contains a brief introduction to the subject followed by papers on applications in biology, urban economics, history

and philosophy. AM

Differential Equations, T(15-16: 1), S, L. Differential Equations. A.N. Tikhonov, A.B. Vasil'eva, A.G. Sveshnikov. Transl: A.B. Sossinskij. Springer-Verlag, 1985, viii + 238 pp, $38 (P). [ISBN: 0-387-13002-0] General theory (existence and uniqueness theorems, dependence of solutions on initial values and parameters, Picard's Method, Contraction Mapping Theorem), linear differential equations, boundary value problems, stability theory, numerical methods, asymptotic solutions, first-order partial differential equations. No exercises. Excellent exposition. DA

Partial Differential Equations, S(18), P, L. Partial Differential Equations: New Methods for their Treatment and Solution. Richard Bellman, George Adomian. Math. & Its Applic. D Reidel, 1985, xvii + 290 pp, $49. [ISBN: 90-277-1681-1] Includes new methods which lead to effective numerical algorithms, new methods of obtaining Green's functions for partial differential operators, and new material on solving nonlinear equations by Adomian's decomposition methodology. Considers many types of problems. DA

Partial Differential Equations, P. The Analysis of Linear Partial Differential Operators. Lars Hörmander. Springer-Verlag, 1985. III: Pseudo-Differential Operators, viii + 525 pp, $48.50 [ISBN: 0-387-13828-5]; IV: Fourier Integral Operators, vii + 352 pp, $45. [ISBN: 0-387-13829-3] A systematic development, with applications, of the theory of pseudo-differential operators, Lagrangian distributions and Fourier integral operators, and symplectic geometry as tools in the study of linear differential equations with variable coefficients. AO

Partial Differential Equations, S*(17-18), P*. The Boundary Value Problems of Mathematical Physics. O.A. Ladyzhenskaya. Transl: Jack Lohwater. Appl. Math. Sci., V. 49. Springer-Verlag, 1985, xxx + 322 pp, $58. [ISBN: 0-387-90989-3] Translated from the original 1973 Russian edition Craevie Zadachi Matematicheskoi. In the present edition, "Supplements and Problems" have been added at the ends of the six chapters to illustrate the possibilities of the book's methods and to provide topics for independent work by students. Systematic exposition of problems of solvability of linear boundary and initial boundary value problems from partial differential equations (mainly of order two) with variable coefficients. JK

Functional Analysis, S(18), P. Banach Lattices and Operators. Hans-Ulrich Schwarz. Teubner-Texte zur Math., B. 71. BG Teubner, 1984, 208 pp, 19,50 M (P). Includes three chapters: vector lattices, Banach lattices, and bounded operators in Banach lattices. Bibliography. JS

Functional Analysis, T(18: 2), S, P. Nonlinear Functional Analysis. Klaus Deimling. Springer-Verlag, 1985, xiv + 450 pp, $39. [ISBN: 0-387-13928-1] Assuming a fairly modest background in analysis, the author begins with topological degree for maps on R^n then generalizes to infinite dimensional spaces. A number of nonlinear methods and applications are considered including implicit functions, fixed points, approximate solutions, multi-valued maps, extremal problems, and bifurcation. Exercises, bibliography. JS

Functional Analysis, S(18), P. The Fixed Point Index and Some Applications. Roger D. Nussbaum. Pr U Montreal, 1985, 145 pp, $18 (P). [ISBN: 2-7606-0711-9] Intended as "propaganda" for the usefulness of the fixed point index of a mapping in a topological space, the author generalizes the classical Leray-Schauder degree, beginning with the theory in R^n and then extending it. Treats the Krein-Rutman theorem, the "mod p" theorem, and applications to bifurcation theory and differential equations. References. JS

Functional Analysis, S(18), P. Rings of Continuous Functions. Ed: Charles E. Aull. Lect. Notes in Pure & Appl. Math.,V. 95. Dekker, 1985, x + 318 pp, $65 (P). [ISBN: 0-8247-7144-3] With some additions and deletions the contents are essentially the proceedings of a special session on rings of continuous functions at the AMS winter meeting in 1982. Additional section on problems raised by individual papers. JS

Functional Analysis, P. Operators and Function Theory. Ed: S.C. Power. NATO ASI Ser., C: V. 153. D Reidel, 1985, xvi + 383 pp, $49. [ISBN: 90-277-2008-8] Proceedings of the NATO Institute on Operators and Function Theory held at the University of Lancaster (U.K.) in July, 1984. Major papers are included with a list of participants and a complete program. JS

Functional Analysis, S(18), P. Lecture Notes in Mathematics-1120: Perturbations of Banach Algebras. Krzysztof Jarosz. Springer-Verlag, 1985, 118 pp, $9.80 (P). [ISBN: 0-387-15218-0] An introduction to a nascent theory concerning small deformations of the algebraic structure and almost isometric invariants in the theory of Banach algebras. Based on the author's most recent paper and unpublished results, together with some of the work by R. Rochberg and B.E. Johnson. MU

Functional Analysis, S(18), P. Almost-periodic Functions in Abstract Spaces. S. Zaidman. Res. Notes in Math., V. 126. Pitman, 1985, 133 pp, $14.95 (P). [ISBN: 0-273-08661-8] Presentation of some recent results on almost-periodic functions with values (usually) in a Banach or Hilbert space. Brief introductory theory is followed by discussion of various families of functions including almost-automorphic and asymptotically almost-periodic functions. References, index. JS

Analysis, T(16-17: 2), L. Equations of Mathematical Physics. V.S. Vladimirov. Transl: Eugene Yankovsky. MIR, 1984, 464 pp, $10.95. Examines the well-posed boundary value problems of classical mathematical physics, using the concept of generalized solution (which is based on those of generalized derivative and function). Assumes some background in vector analysis, elementary complex vari-

ables and Fourier series; provides necessary Lebesgue theory and functional analysis. For students of mathematics, physics, engineering. Sequel to the author's 1971 book of the same title (TR, May 1971). DFA

Analysis, S(16). Studies on Divergent Series and Summability and the Asymptotic Developments of Functions Defined by Maclaurin Series. Walter B. Ford. Chelsea, x + 342 pp, $19.50. Combined reprints of works originally published in 1916 and in 1936, respectively. Dated in style and content but good for seeing how mathematics used to be written. The material was current when the books were first produced and Ford was active in its development. JK

Analysis, T(16-17: 2). Holomorphy and Calculus in Normed Spaces. Soo Bong Chae. Pure & Appl. Math., V. 92. Dekker, 1985, xii + 421 pp, $65. [ISBN: 0-8247-7231-8] In three parts: Part I contains material covering calculus in normed spaces; Part II extends results from complex analysis to holomorphic mappings between normed spaces; Part III treats locally convex topologies on spaces of holomorphic mappings. A valuable compilation of results in this field. Unfortunately the presentation is quite formal, and the price quite steep. TAV

Algebraic Geometry, P. Lecture Notes in Mathematics-1124: Algebraic Geometry, Sitges (Barcelona) 1983. Ed: E. Casas-Alvero, G.E. Welters, S. Xambó-Descamps. Springer-Verlag, 1985, xi + 416 pp, $25.80 (P). [ISBN: 0-387-15232-6] Thirteen papers from a conference in Sitges (Barcelona), Spain, in October 1983 on "classical" algebraic geometry--i.e., curves and surfaces in projective space. Several are fairly long, with complete proofs, expository sections, and numerous examples. BC

Differential Geometry, T*(17-18), S*. Differential Manifolds and Theoretical Physics. W.D. Curtis, F.R. Miller. Pure & Appl. Math., V. 116. Academic Pr, 1985, xix + 394 pp, $69. [ISBN: 0-12-200230-X] Introduces vector fields, geodesics, differential forms, integration, Frobenius theorem, Lie groups, and principal fiber bundles. Uses physics to motivate these topics and provide applications, e.g., applies the theory of Lie groups to rigid body motion to illustrate the connection between symmetry and momentum conservation in classical mechanics. AM

Geometry, P*. Modern Geometry--Methods and Applications, Part II: The Geometry and Topology of Manifolds. B.A. Dubrovin, A.T. Fomenko, S.P. Novikov. Transl: Robert G. Burns. Grad. Texts in Math., V. 104. Springer-Verlag, 1985, xv + 430 pp, $54. [ISBN: 0-387-96162-3] The second volume of a three-volume introduction to modern geometry. Outcome of a reworking and extensive elaboration of lecture notes for an archetypal course at Moscow State University. At a somewhat higher level than Part I. Guiding principle is "minimal abstractness of exposition, giving preference...to significant examples over...general theorems." Homotopy groups, fibre bundles, dynamical systems, and foliations. Emphasis on applications to other areas of mathematics and theoretical physics. JK

Geometry, T??(13: 1). Synthetic Geometry. B.M. Saler (30 Melva Crescent, Agincourt, Ontario, Canada M1V 1A3), 1984, 205 pp, $20 (P). An unimpressive photocopy presentation of synthetic projective geometry without any obvious purpose; exercises but no index. JNC

Geometry, S(14), L. Analytic Geometry. V.A. Ilyin, E.G. Poznyak. Transl: Irene Aleksanova. MIR, 1984, 232 pp, $7.95. Based on lectures at the physics department of Moscow University; chapters cover systems of coordinates, determinants, vector algebra, transformations of the plane, equations of lines, planes, surfaces, and curves. JNC

Geometry, T??(13: 1). Elementary Plane Geometry. R. David Gustafson, Peter D. Frisk. Wiley, 1985, x + 358 pp, $26.95. [ISBN: 0-471-89047-2] Intended as a "quick review," it reinforces misconceptions of many high school presentations (e.g., glossing over distinctions between line segments and measures, giving all proofs in two-column format). Among "major changes" in this Second Edition (First Edition, TR, April 1973) are the numbering of sections and exercise sets and inclusion of reasons for every statement in proofs. JNC

Algebraic Topology, S(18), P. Lecture Notes in Mathematics-1116: Hochzusammenhängende Mannigfaltigkeiten und ihre Ränder. Stephan Stolz. Springer-Verlag, 1985, xxiii + 134 pp, $12 (P). [ISBN: 0-387-15209-1]

Algebraic Topology, T(16-18: 1, 2). Elements of Algebraic Topology. James R. Munkres. Addison-Wesley, 1984, ix + 454 pp, $35.95. [ISBN: 0-201-04586-9] Assumes a solid undergraduate course in each of modern algebra and point set topology. Presents simplicial homology, goes through enough homological algebra to present the Eilenberg-Steenrod Axioms (category theory is minimal and optional), and then develops singular homology theory. Concludes with some homological algebra and applications to duality in manifolds. Purposely omits homotopy theory. The result is a beautifully written book that bridges the gap between a classical geometrical approach and a slick algebraic approach. Good exercises and index; short and to the point bibliography. JAS

Operations Research, T(16-17: 1, 2), L.** Fundamentals of Queueing Theory, Second Edition. Donald Gross, Carl M. Harris. Ser. in Prob. & Math. Stat. Wiley, 1985, xiii + 587 pp, $42.95. [ISBN: 0-471-89067-7] A thorough and thoughtful revision of an already good book (First Edition, TR, December 1974). New material on queueing networks and approximations highlight the changes. Computer applications reflect new simulation languages. A satisfying textbook. TAV

Optimization, S(16-17), L. Problems and Exercises in the Calculus of Variations. M.L. Krasnov, G.I. Makarenko, A.I. Kiselev. Transl: George Yankovsky. MIR, 1984, 222 pp, $7.95 (P). Basic definitions and theorems. Well-exampled, with over 200 exercises-problems. Hints and/or answers. Specially

designed for readers using Differential Equations and the Calculus of Variations by Elsgolts (MIR, 1970). Extrema of functions of several variables, and direct methods in the calculus of variations. Excellent problem source. (Reprint of 1975 original edition, TR, February 1976.) JK

Dynamical Systems, S*(16-17), P, L. Dynamics--The Geometry of Behavior, Part 3: Global Behavior. Ralph H. Abraham, Christopher D. Shaw. Visual Math. Lib. Aerial Pr, 1985, xi + 123 pp, $26 (P). [ISBN: 0-942344-03-4] Structural stability concerns those features of a dynamical system that persist after a perturbation. Generic properties characterize the features of typical dynamical systems. This book (Part 1, TR, August-September 1983; Part 2, TR, March 1984) presents visual, pictorial discussions explaining these concepts. Develops definitions, theorems, proofs primarily through illustrations. Accessible to undergraduates. Refers to preceding volumes. AM

Control Theory, T(18: 1), S, P. Control System Synthesis: A Factorization Approach. M. Vidyasagar. Ser. in Signal Processing, Optimization, & Control, V. 7. MIT Pr, 1985, 436 pp, $35. [ISBN: 0-262-22027-X] Designed for a second-level text and research monograph, the author assumes a background including a first-year graduate course in linear system theory and sufficient algebra background to include basic ring theory and topological rings. His objective is to present the "factorization" approach as a powerful tool relating to various areas of control theory including stabilization, filtering, and robustness. Some exercises, references, three appendices on algebra. JS

Control Theory, P. Lecture Notes in Mathematics-1119: Recent Mathematical Methods in Dynamic Programming. Ed: I. Capuzzo Dolcetta, W.H. Fleming, T. Zolezzi. Springer-Verlag, 1985, 202 pp, $14.40 (P). [ISBN: 0-387-15217-2] Contains ten papers presented at Università de Roma, March 26-28, 1984. The papers cover analytical, numerical and applied aspects of control as well as related topics in partial differential equations, functional analysis and stochastic processes. TAV

Control Theory, P. Stochastic Modelling and Control. M.H.A. Davis, R.B. Vinter. Mono. on Stat. & Appl. Prob. Chapman & Hall, 1984, xii + 393 pp, $39.95. [ISBN: 0-412-16200-8] The emphasis is on modelling systems using the nature of input and output data. Linear systems, filters, parameter estimation and asymptotic analysis are emphasized. The treatment is quite rigorous. TAV

Systems Theory, P, L. Nonlinear System Theory. John L. Casti. Math. in Sci. & Eng., V. 175. Academic Pr, 1985, xi + 261 pp, $45. [ISBN: 0-12-163452-3] An inviting, advanced survey of finite dimensional deterministic nonlinear system theory, replete with concrete examples and careful references. Emphasizes (and presents) algebraic and geometric tools rather than functional-analytic methods; includes a survey of modern linear system theory in abstract form (e.g., as theorems about torsion modules and invariant factors). Then covers controllability, observability and stability for nonlinear systems in similar terms, using ideas for algebraic and differential geometry. LAS

Probability, P. Probabilistic Methods in the Mechanics of Solids and Structures. Ed: S. Eggwertz, N.C. Lind. Intern. Union of Theor. & Appl. Mech. Springer-Verlag, 1984, xxvi + 610 pp, $51. [ISBN: 0-387-15087-0] Proceedings of an international symposium in honor of Waloddi Weibull, Stockholm, June 1984. Following Weibull's work, there are papers on extreme value theory, probabilistic failure models, reliability analysis, and structural design. TAV

Statistics, P. Inférence Statistique Analyse des Données Sous des Plans D'Échantillonnage Complexes. Carl E. Särndal. Pr U Montreal, 1984, 123 pp, $17 (P). [ISBN: 2-7606-0676-7]

Statistics, P, L*. Herbert Robbins: Selected Papers. Ed: T.L. Lai, D. Siegmund. Springer-Verlag, 1985, xli + 518 pp, $36. [ISBN: 0-387-96137-2] A selection of nearly half of Robbins' 133 publications arranged in three groups: empirical Bayes methodology, sequential experimentation, and probability and inference. Begins with a complete list of Robbins' publications and an interview with Robbins reprinted from the College Mathematics Journal. LAS

Statistics, S(13-15), L*. Say It With Figures, Sixth Edition. Hans Zeisel. Harper & Row, 1985, xvi + 272 pp, $14.50; $7.95 (P). [ISBN: 0-06-181982-4; 0-06-131994-5] First published in 1947, Say It With Figures is a classic presentation of practical issues involved with the use of data in social science: presentation of data, experimental vs. non-experimental results, inference of cause and effect. This latest edition adds a very helpful chapter on regression analysis. LAS

Statistics, S*(14-18), P, L*. Encyclopedia of Statistical Sciences, Volume 5: Lindberg Condition to Multitrait-Multimethod Matrices. Ed: Samuel Kotz, Norman L. Johnson. Wiley, 1985, ix + 741 pp, $85. [ISBN: 0-471-05552-2] Fifth volume of a planned eight-volume set for the non-specialist. (See TR, October 1982 of Series and Volume 1.) RSK

Statistics, P*. The Collected Works of John W. Tukey, Volume II: Time Series: 1965-1984. Ed: David R. Brillinger. Wadsworth, 1984, lxvii + 582 pp, $37.95. [ISBN: 0-534-03304-0] Second volume of a projected series covering Tukey's contributions to statistics (see TR, May 1985 of Series and Volume 1). The first two volumes contain all his important papers on time series, including many previously unpublished. RSK

Computer Literacy, S, P. A UNIX Primer. Ann Nicols Lomuto, Nico Lomuto. Prentice-Hall, 1983, xvi + 239 pp, $20 (P). [ISBN: 0-13-937731-X] Primarily an introduction to text editing on UNIX, using "ed" and "nroff". Devotes much attention to writing "nroff" macros and illustrating fancy editing and formatting problems. Pipes and filters are introduced only at the end; shell features are barely mentioned, programming not at all. LAS

Computer Literacy, T?(13: 1). Computers Today, Second Edition. Donald H. Sanders. McGraw-Hill, 1985, xxiii + 648 pp, $23.95 [ISBN: 0-07-054701-7]; Test Bank, 272 pp, (P) [ISBN: 0-07-054703-3]; Instructor's Manual, iii + 225 pp, (P). [ISBN: 0-07-054702-5] Lots of pretty pictures, lots of description of what everybody is doing, lots of gee-whiz stuff and very little intellectual substance. The author does comment on the rate at which such books get out of date, and falls prey to his own prediction, e.g., his data on the Cray 2 are wrong, and he gives support to many popular rumors about standards. Many high-school students might find this interesting; many would giggle and find it boring. JAS

Computer Literacy, T(13-14: 1), L. The Information Technology Revolution. Ed: Tom Forester. MIT Pr, 1985, xvii + 674 pp, $14.95 (P). [ISBN: 0-262-56033-X] A sequel to the 1980 collection The Microcomputer Revolution intended to help bridge C.P. Snow's gulf of mutual incomprehension between the two cultures of science and liberal arts. 48 reprints (by scientists and journalists) deal with the impact of computing on work, education, science, and society. Each of the 14 chapters ends with a brief bibliography for further reading. LAS

Computer Literacy, S(13). Apple Macintosh Encyclopedia. Gary Phillips, Donald J. Scellato. Chapman & Hall, 1984, 328 pp, $19.95 (P). [ISBN: 0-412-00671-5] An alphabetic glossary of common terms encountered in reading Macintosh literature, from "accented characters" to "user friendly." Little substance. LAS

Computer Literacy, T(13-16: 1). Computer Annual: An Introduction to Information Systems 1985-1986. Robert H. Blissmer. Ser. in Comp. & Inform. Proc. Syst. for Business. Wiley, 1985, xiv + 487 pp, $20.95 (P). [ISBN: 0-471-81106-8] Covers organization and development of computer and information systems, from personal computers to mainframes. Numerous photos and useful drawings highlight readable and organized chapters; each begins with a detailed outline and ends with complete review exercises. RD

Computer Literacy, L. Computer Culture: The Scientific, Intellectual, and Social Impact of the Computer. Ed: Heinz R. Pagels. Annals, V. 426. NY Academy of Sci, x + 288 pp, $66 (P). [ISBN: 0-89766-245-8] Papers from an April 1983 Science Week symposium sponsored by the New York Academy of Sciences in which leaders of the computer revolution engage non-specialists in thinking through the societal implications of computer technology. Includes record of panel discussions and post-lecture questions and answers. The price of $.25 per page is quite high. LAS

Computer Literacy, S(13). Computers: Understanding and Using Them, A Hands-On Approach. R.L. Richardson, Janet Gilchrist. Gorsuch Scarisbrick, x + 211 pp, $18 (P). [ISBN: 0-89787-409-9] Gives beginner "working knowledge of computer and software packages." Progresses through word processing, data bases and spreadsheets, plus chapters on impact of the computer and on BASIC programming. Each chapter includes either discussion questions or software/programming examples. RD

Computer Literacy. The Osborne/McGraw-Hill CP/M-86 User's Guide. Jonathan Sachs. Osborne McGraw-Hill, 1985, viii + 568 pp, $18.95 (P). [ISBN: 0-88134-143-6] Very much a user-oriented manual for computers using some version of CP/M-86, MP/M-86 and such. Covers cleaning disk drives, "communicating with other computers," and general information rather than operating system internals. JAS

Computer Programming, T(13: 1, 2), S. PASCAL Plus Data Structures, Algorithms, and Advanced Programming. Nell Dale, Susan C. Lilly. DC Heath, 1985, xix + 635 pp, $26.95 (P). [ISBN: 0-669-07239-7] A text for the second half of the standard introduction to computer science, emphasizing advanced topics and substantial applications for students who have completed a first course in Pascal: stacks, queues, linked lists, pointers, recursion, trees, searching, hashing. Extensive appendices provide reference details on Pascal, answers to exercises, glossary, and a collection of substantial programming assignments. LAS

Computer Programming. The Master Handbook of High-Level Microcomputer Languages. Charles F. Taylor. TAB Books, 1984, viii + 359 pp, $15.50 (P). [ISBN: 0-8306-1733-7] A description of the major features of BASIC, C, COBOL, Forth, FORTRAN, LISP, Logo, Modula-2, Pascal, and PILOT as found on Apple, IBM, or generic CP/M computers. JAS

Computer Programming, S(13-15), P*, L. Handbook of BASIC for the IBM PC, Revised and Expanded. David I. Schneider. Prentice-Hall, 1985, ix + 579 pp. [ISBN: 0-89303-510-6] An innovative, effective and informative work--a reference worth reading--by an experienced mathematician author. Each command is explained thoroughly in literate and mathematically sound terms, frequently with apt illustrations for the mathematically-minded user. Includes good illustrations of each command. LAS

Computer Programming, T(14-16: 1), S, P, L. Programming in Prolog, Second Edition. W.F. Clocksin, C.S. Mellish. Springer-Verlag, 1984, xv + 297 pp, $17.95 (P). [ISBN: 0-387-11046-1] Corrected and improved new edition of the 1981 original edition. Prolog is a non-procedural language used for logic programming, with applications in symbolic algebra, relational databases, and artificial intelligence. LAS

Computer Programming, S, P, L. 8087 Applications and Programming for the IBM PC and Other PCs. Richard Startz. Prentice-Hall, 1983, xii + 276 pp, (P). [ISBN: 0-89303-420-7] A clear introduction to the benefits (and incompatibilities) of the Intel 8087 chip, the fast arithmetic coprocessor designed to work alongside the IBM PC's 8088 CPU. Explains what's involved in software conversion, discusses the encoding of numbers and other features of 8087 architecture, and offers several dozen BASIC and assembler programs (available on disk) for linear algebra and non-linear functions. LAS

Computer Programming, T(13: 1). Structured BASIC. Richard M. Jones. Allyn & Bacon, 1985, xii + 419 pp, $25 (P). [ISBN: 0-205-08271-8] For introduction to BASIC programming or microcomputer applications. Primarily Microsoft BASIC, with discussion of other common versions. Includes discussion of structured algorithm design, program development tools, CP/M, MS-DOS, dBASEII, systems software, language, peripherals. Stresses menu driven programs, models and simulations, reports, cursor control. RM

Computer Programming, T?(13-14), S. Problem Solving Using UCSD Pascal, Second Edition. Kenneth L. Bowles, Stephen D. Franklin, Dennis J. Volper. Springer-Verlag, 1984, xi + 340 pp, $17.95 (P). [ISBN: 0-387-90822-6] Disappointing considering the authors, the publisher, and that it is a rewrite of Microcomputer Problem Solving Using Pascal (1977). The first three places the reviewer looked he found errors: the use of '/' instead of 'div', a ';' instead of ',' in a syntax chart, and a description of a non-standard built-in procedure that indicated call-by-variable when call-by-value was correct. Little on general problem solving; much on using Turtle Graphics and strings. RWN

Computer Programming, P. The Applesoft BASIC Primer For the Apple II Plus, IIe, and IIc. Jo Lynne Talbott Jones. Comp. in Educ. Ser. Computer Science Pr, 1985, vi + 74 pp, $9.95 (P). [ISBN: 0-88175-047-6] Very basic guide for any adult helping to introduce computers and programming to children. Includes exercise and "teacher" guide. Questionable use of BASIC as initial programming experience for children. RD

Computer Programming, P. Logoworlds. Rachelle S. Heller, C. Dianne Martin, June L. Wright. Comp. in Educ. Ser. Computer Science Pr, 1985, xiii + 301 pp, $19.95 (P). [ISBN: 0-88175-031-X] Intended as a guide for teachers implementing LOGO in the K-8 classroom. Written for Apple Logo. Comprehensive and sequential; includes LOGO primitives, basic and advanced graphics, list processing, and numerical operations. RD

Software Systems, S*, P*, L*. The UNIX Programming Environment. Brian W. Kernighan, Rob Pike. Prentice-Hall, 1984, x + 357 pp, $26.95. [ISBN: 0-13-937699-2] A thorough, no-nonsense systematic survey of UNIX tools, from fundamentals to programming, written by people who developed much of UNIX. The first several chapters are devoted to the shell: commands, filters, programming. The final two chapters on program development and document preparation use as an extended example the language "hoc", which emulates a high order calculator. An excellent in-depth introduction to UNIX for those willing to work and learn. LAS

Computer Science, P. Logics and Models of Concurrent Systems. Ed: Krzsztof R. Apt. NATO ASI Ser. F.: Comp. & Syst. Sci., V. 13. Springer-Verlag, 1985, viii + 498 pp, $53.90. [ISBN: 0-387-15181-8] Proceedings of a NATO Advanced Course, October 1984. Papers discuss temporal logic, syntax directed verification, CSP. RM

Computer Science, T(13: 1, 2). Introduction to Computer Science, Third Edition. Neill Graham. West, 1985, xiii + 894 pp. [ISBN: 0-314-85240-9] This text, for a first computer science course (not programming), is designed for a mathematically unsophisticated student (only one year of algebra). For this reason it may be more suited to a high-school Advanced Placement course, than for a college-level course; the presentation seems clear but dilute. Pascal is introduced and developed with supplements starting with Chapter 6; nevertheless, the book is relatively language-independent. JAS

Computer Science, P. Lecture Notes in Computer Science-189: Paragon. Mark Steven Sherman. Springer-Velag, 1985, xi + 364 pp, $20.50 (P). [ISBN: 0-387-15212-1] Author's Ph.D. thesis on language facilities to support semi-automatic specification, implementation, and selection of data abstractions, illustrated via the language Paragon. Classes give type hierarchies which are used to refine specifications into implementations, with compile-time selection of (possibly multiple) implementations. RM

Computer Science, P. Lecture Notes in Computer Science-191: A Survey of Verification Techniques for Parallel Programs. Howard Barringer. Springer-Verlag, 1985, vi + 115 pp, $11.20 (P). [ISBN: 0-387-15239-3] Survey of current trends and principles of specificiation and verification techniques for parallel and distributed programs. Techniques divided between those for shared variable parallelism and for message-based (e.g., Ada tasks, CSP) parallelism. RM

Computer Science, P. Lecture Notes in Computer Science-185 & 186. Ed: Hartmut Ehrig, et al. Springer-Verlag, 1985. Mathematical Foundations of Software Development, Volume 1: Colloquium on Trees in Algebra and Programming, xiii + 418 pp, $22.80 (P) [ISBN: 0-387-15198-2]; Formal Methods and Software Development, Volume 2: Colloquium on Software Engineering (CSE), xiv + 455 pp, $25.10 (P). [ISBN: 0-387-15199-0] Proceedings of the Joint Conference on the Theory and Practice of Software Development held in Berlin, West Germany, March 1985. Contains about 50 invited papers addressing the issue of the use of formal mathematics (especially in the area of formal language theory and semantics) in the area of software design and development. MS

Computer Science, S(16), P. Verification and Validation of Real-Time Software. Ed: W.J. Quirk. Springer-Verlag, 1985, xi + 245 pp, $29.50. [ISBN: 0-387-15102-8] Real-time software are programs which run in a time-critical environment and control on-going physical processes. Examples might include flight navigation systems or manufacturing control programs. It is critically important that these programs be thoroughly debugged, as failures can cause not simply irritations but physical or personal damage as well. This book describes techniques for validating the correctness of these special real-time programs. The techniques include empirical testing, simulation, statistical

sampling, and formal verification. MS

Computer Science, S(16), P. Heuristic Reasoning about Uncertainty: An Artificial Intelligence Approach. Paul R. Cohen. Research Notes in Artif. Intell., V. 2. Pitman, 1985, 204 pp, $19.50 (P). [ISBN: 0-273-08667-7] Describes a technique, called the model of endorsement, for manipulating and reasoning about uncertain information. This theory, based on the principles of artificial intelligence, is offered as an alternative to other techniques for dealing with uncertainty, including probability theory and fuzzy logic. The model has been implemented as a computer program termed SOLOMON. MS

Computer Science, S(16-18), P. Computer Communications. Ed: B. Gopinath. Proc. of Symp. in Appl. Math., V. 31. AMS, 1985, ix + 124 pp, $27. [ISBN: 0-8218-0082-5] Papers from an AMS Short Course given at Denver in January 1983. Emphasizes by example the diverse mathematics involved in such things as the complexity of VLSI circuits, concurrent processes, and modelling information flow in a communications network. LAS

Computer Science, P. Lecture Notes in Computer Science-188: Advances in Petri Nets 1984. Ed: G. Rozenberg. Springer-Verlag, 1985, vii + 467 pp, $25.10 (P). [ISBN: 0-387-15204-0] Twenty-five papers, representing the most significant results from the past two workshops on Applications and Theory of Petri Nets (Toulouse 1983, Aarhus 1984). LCL

Computer Science, T(14-15: 1, 2), L. Software Development in Pascal. Sartaj Sahni. Camelot, 1985, x + 690 pp, $30. [ISBN: 0-942450-01-9] An introduction to all aspects of the software development task. Topics include problem specification, human interface design, modularization, program specificiation, program aesthetics, stepwise refinement, testing, documentation, verification, performance analysis and measurement, data structures, and algorithm design methods. AO

Computer Science, P. Algorithmically Specialized Parallel Computers. Ed: Lawrence Snyder, et al. Academic Pr, 1985, xiii + 252 pp, $26.50. [ISBN: 0-12-654130-2] Papers from a 1982 workshop at Purdue on specialized architectures designed for efficient execution of certain structural classes of algorithms. Discusses trade-offs between architectural optimization for specialized performance and loss of generality. Use of VLSI, architectures for signal, speech, and image processing, numerical computations. RM

Computer Science, P. Information Management for Engineering Design. Randy H. Katz. Surveys in Comp. Sci. Springer-Verlag, 1985, 93 pp, $22.50. [ISBN: 0-387-15130-3] Discusses issues of managing information about large design projects (engineering design applications) and limitations of current CAD systems, given the difficulty of applying database technology to the problems encountered by CAD systems. Synthesis and analysis tools, design data model, object model. RM

Computer Science, T(16-17). Data Base Management Systems, Second Edition. Alfonso F. Cardenas. Allyn & Bacon, 1984, xix + 745 pp, $37.44. [ISBN: 0-205-08191-6] Extension of the 1979 edition (TR, October 1980) on the technology, practice, and implementation of database systems. Concepts, schemas, architectures, relational and network (CODASYL) approaches, normalization, high-level file organization. Discusses current systems (IMS, TOTAL, System 2000, SQL, QBE); includes new chapters on database machines, distributed databases. RM

Applications, S(16-18), P, L. Learning the Art of Mathematical Modelling. Mark Cross, A.O. Moscardini. Ser. in Math. & Its Applic. Halsted Pr, 1985, 155 pp, $34.95; 15.95 (P). [ISBN: 0-470-20168-1] Not a textbook but rather an engaging, informal introduction to the art of mathematical modelling aimed at both the teacher and the student. Several models are presented in detail but the concern is more with general concepts and how to teach them, complete with anecdotal experience and a chapter on computer software. JS

Applications (Artificial Intelligence), P. A Critiquing Approach to Expert Computer Advice: ATTENDING. Perry L. Miller. Pitman, 1984, 112 pp, $19.50 (P). [ISBN: 0-273-08665-0] Describes experimental ATN-based expert system for medical management (pre-operative anesthetic plans) rather than diagnosis. Generates prose critiques of physician's plans, based on explicit heuristic (non-statistic) risk analysis; runs in consultative or tutorial mode. RM

Applications (Artificial Intelligence), S(16-18), P*, L. Searching with Probabilities. Andrew J. Palay. Research Notes in Artif. Intell., V. 3. Pitman, 1985, 192 pp, $19.50 (P). [ISBN: 0-273-08664-2] This seminal work examines the use of probability distribution functions within the domain of game playing algorithms (with special attention to chess). There is evidence to suggest that the probability-based algorithm will eventually surpass the ability of the alpha-beta algorithm. LCL

Applications (Biology), S(15-18), P, L. Trees and Networks in Biological Models. N. MacDonald. Wiley, 1983, ix + 215 pp, $39.95. [ISBN: 0-471-10508-2] A self-contained exposition of applications of graph theory in biology. The first part deals with abstract networks (especially dominance and predation); the second part treats dynamics using ordinary differential equations or difference equations; the third and longest part is concerned with rooted trees. No exercises. LCL

Applications (Biology), P. New Phenology: Elements of Mathematical Forecasting in Ecology. Alexander S. Podolsky. Wiley, 1984, xiii + 504 pp, $64.95. [ISBN: 0-471-86451-X] Presents methods for predicting development times from temperature data. FLW

Applications (Engineering), P*. Variational Methods in Engineering. Ed: C.A. Brebbia. Springer-Verlag, 1985, 538 pp, $89. [ISBN: 0-387-15496-5] Proceedings of the Second International Conference on Variational Methods in Engineering held at the University of Southampton, England, in July 1985. On the formulation of variational or related principles and the solution of the corresponding stationary problem. Over 40 papers including keynote lectures. Photocopied from typewritten manuscripts. JK

Applications (Engineering), S*(13-14), L*. Mathematics in Action. O.G. Sutton. Dover, 1984, viii + 236 pp, $4.95 (P). [ISBN: 0-486-24759-7] Reprint of the 1966 Second Edition. Light but not superficial treatment of applied mathematics illustrated with examples from gunnery, wave theory, mathematics of flight, statistics and meteorology. A bit dated but quite readable. Knowledge of some rudimentary calculus is sufficient for enjoyment. JK

Applications (Physics), P. Quantum Fluctuations. Edward Nelson. Ser. in Physics. Princeton U Pr, 1985, viii + 146 pp, $32; $12.95 (P). [ISBN: 0-691-08378-9; 0-691-08379-7] An introduction to stochastic mechanics (a description of quantum phenomena in classical probabilistic terms). Includes many recent developments. AO

Applications (Physics), P. Lecture Notes in Mathematics-1109: Stochastic Aspects of Classical and Quantum Systems. Ed: S. Albeverio, Ph. Combe, M. Sirugue-Collin. Springer-Verlag, 1985, ix + 227 pp, $14.40 (P). [ISBN: 0-387-13914-1] Proceedings of the Second French-German Encounter in Mathematics and Physics, held in Marseille, France, March 28 to April 1, 1983. JAS

Applications (Physics), P. Transformation Groups Applied to Mathematical Physics. Nail H. Ibragimov. Math. & Its Applic. D Reidel, 1984, xv + 394 pp, $69. [ISBN: 90-277-1847-4] Group theoretical techniques applied to motions in Riemannian spaces, the Huygens principle, evolution equations, and conservation laws. AO

Applications (Physics), P. Lecture Notes in Physics-220: Effective Lagrangians in Quantum Electrodynamics. Walter Dittrich, Martin Reuter. Springer-Verlag, 1985, 244 pp, $14.60 (P). [ISBN: 0-387-15182-6] An introduction to the subject of effective Lagrangians for quantum electrodynamics. Illustrates several important computational techniques not usually covered in standard texts. AO

Applications (Physics), P. Lecture Notes in Physics-218: Ninth International Conference on Numerical Methods in Fluid Dynamics. Ed: Soubbaramayer, J.P. Boujot. Springer-Verlag, 1985, x + 612 pp, $32 (P). [ISBN: 0-387-13917-6] Proceedings of a conference held at the Centre d'Etudes Nucléaires de Saclay in France, June 25-29, 1984. JAS

Applications (Physics), T(13-14: 1, 2). Introduction to Physical Mathematics. P.G. Harper, D.L. Weaire. Cambridge U Pr, 1985, xi + 260 pp, $42.50. [ISBN: 0-521-26278] Largely a rehash of material covered in introductory courses in calculus, linear algebra, and differential equations. Much space is devoted to such elementary concepts as differentiation and integration, leaving barely any space to explain somewhat more advanced topics such as curl and divergence. MU

Applications (Physics), P. Lecture Notes in Physics-216: Applications of Field Theory to Statistical Mechanics. Ed: L. Garrido. Springer-Verlag, 1985, viii + 352 pp, $23.70 (P). [ISBN: 0-387-13911-7] Proceedings of the Sitges Conference on Statistical Mechanics held in Sitges and Barcelona, Spain in June 1984. BH

Applications (Social Science), P. Mathematical Ideas and Sociological Theory: Current State and Prospects. Ed: Thomas J. Fararo. Gordon & Breach, 1984, 405 pp, $24.50 (P). [ISBN: 0-677-16635-4] Seven articles constituting a special issue of the Journal of Mathematical Sociology, each reflecting on the present state and future prospects of mathematical theory applied to sociology. LAS

Applications (Social Science), S(15-17), P, L*. Superpower Games: Applying Game Theory to Superpower Conflict. Steven J. Brams. Yale U Pr, 1985, xvi + 176 pp, $6.95 (P); $22.50. [ISBN: 0-300-03364-8; 0-300-03323-0] An innovative and important analysis of deterrence, arms race, and verification modelled on such games as Chicken and Prisoners' Dilemma. Brams suggests that optimal strategy often involves probabilistic threats, mutual predictability, and other sophisticated, counterintuitive strategies. LAS

Reviewers

DFA: David F. Appleyard, Carleton; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; RD: Roger Day, St. Olaf; JD-B: John Dyer-Bennet, Carleton; SG: Steven Galovich, Carleton; BH: Bruce Hanson, St. Olaf; PH: Paul Humke, St. Olaf; KK: Kenneth Kaminsky, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; AM: Alan Magnuson, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; LN: Linda Ness, Carleton; AO: Arnold Ostebee, St. Olaf; MS: Michael Schneider, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MT: Michael Tveite, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

AN ELEMENTARY PROOF OF STIRLING'S FORMULA

P. DIACONIS[1]
Statistics Department, Stanford University, Stanford, CA 94305

D. FREEDMAN[2]
Statistics Department, University of California, Berkeley, CA 94720

1. Introduction. Stirling's formula is

$$\Gamma(\alpha) \approx \left(\frac{\alpha-1}{e}\right)^{\alpha-1}\sqrt{2\pi(\alpha-1)} \tag{1}$$

as $\alpha \to \infty$, in the sense that the ratio of the two sides tends to 1. By definition,

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\,dx \quad \text{for } \alpha > 0. \tag{2}$$

Thus, for positive integers n, $\Gamma(n) = (n-1)!$.

The object of this expository note is to give a short but complete version of Laplace's argument for (1). This will be done in the next section, but here is the idea. Write $\exp(x) = e^x$ and $\beta = \alpha - 1$. Then

$$\Gamma(\alpha) = \int_0^\infty \exp\{\Psi_\beta(x)\}\,dx, \tag{3}$$

where

$$\Psi_\beta(x) = \beta\log x - x. \tag{4}$$

Now Ψ_β has its maximum at $x = \beta$, and

$$\Psi_\beta(\beta + y) \doteq \beta\log\beta - \beta - \frac{1}{2\beta}y^2. \tag{5}$$

Here, $\doteq$ means "nearly equal," and is used informally. So

$$\begin{aligned}\Gamma(\alpha) &\doteq \exp(\beta\log\beta - \beta)\cdot\int_{-\infty}^\infty \exp\left(-\frac{1}{2\beta}y^2\right)dy \\ &= \left(\frac{\beta}{e}\right)^\beta\sqrt{2\pi\beta},\end{aligned} \tag{6}$$

because

$$\int_{-\infty}^\infty \exp\left(-\frac{1}{2}z^2\right)dz = \sqrt{2\pi}. \tag{7}$$

2. The argument. Recall Ψ_β from (4). The rigorous version of (5) is the following identity:

$$\Psi_\beta(\beta + y) = \beta\log\beta - \beta - \beta g(y/\beta), \tag{8}$$

where

$$g(v) = v - \log(1 + v). \tag{9}$$

Substitute (8) into (3) and change variables:

$$\Gamma(\alpha) = \left(\frac{\beta}{e}\right)^\beta\int_{-\beta}^\infty \exp\{-\beta g(y/\beta)\}\,dy. \tag{10}$$

Change variables again, putting $y = \sqrt{\beta}\,z$, getting

---

[1] Research partially supported by National Science Foundation Grant MCS80-24649.

[2] Research partially supported by National Science Foundation Grant MCS83-01812.

$$\Gamma(\alpha) = \left(\frac{\beta}{e}\right)^{\beta}\sqrt{2\pi\beta}\,\Gamma_1(\beta), \tag{11}$$

where

$$\Gamma_1(\beta) = \frac{1}{\sqrt{2\pi}}\int_{-\sqrt{\beta}}^{\infty}\exp\{-\beta g(z/\sqrt{\beta})\}\,dz. \tag{12}$$

Our proof of Stirling's formula is reduced to the following.

LEMMA.

$$\lim_{\beta\to\infty}\Gamma_1(\beta) = 1.$$

Proof. Fix L large but finite. Then $\Gamma_1(\beta) = \Gamma_L(\beta) + \tau_L(\beta)$, where

$$\begin{aligned}\Gamma_L(\beta) &= \frac{1}{\sqrt{2\pi}}\int_{-L}^{L}\exp\{-\beta g(z/\sqrt{\beta})\}\,dz \\ &\to \frac{1}{\sqrt{2\pi}}\int_{-L}^{L}\exp\left\{-\frac{1}{2}z^2\right\}dz \quad \text{as } \beta\to\infty,\end{aligned} \tag{13}$$

because

$$\beta g(z/\sqrt{\beta}) \to \frac{1}{2}z^2 \quad \text{as } \beta\to\infty, \text{ uniformly for } z\in[-L,L]. \tag{14}$$

Relationship (14) holds because for v small enough,

$$(1-\varepsilon)\frac{1}{2}v^2 < g(v) < (1+\varepsilon)\frac{1}{2}v^2. \tag{15}$$

It remains to estimate $\tau_L(\beta)$. We take the upper tail first. Abbreviate $h(z) = \beta g(z/\sqrt{\beta})$; the dependence of h on β is implicit. Then

$$\begin{aligned}\int_L^{\infty}\exp\{-h(z)\}\,dz &\leqslant \frac{1}{h'(L)}\int_L^{\infty}h'(z)\exp\{-h(z)\}\,dz \\ &= \frac{1}{h'(L)}\exp\{-h(L)\} \\ &\to \frac{1}{L}\exp\left\{-\frac{1}{2}L^2\right\} \quad \text{as } \beta\to\infty;\end{aligned} \tag{16}$$

the first inequality holds because $h'(z) = \sqrt{\beta}z/(\sqrt{\beta}+z)$ is increasing in $z>0$. The lower tail is similar, so for β large,

$$\tau_L(\beta) \leqslant \frac{2}{\sqrt{2\pi}}\frac{1}{L}\exp\left\{-\frac{1}{2}L^2\right\} + \varepsilon, \tag{17}$$

which completes the proof.

The present approximation comes out in powers of $\beta = \alpha - 1$; if powers of α are preferred, a preliminary change of variables can be made ($x = e^u$), so

$$\Gamma(\alpha) = \int_{-\infty}^{\infty}u^{\alpha}\exp(-e^u)\,du. \tag{18}$$

A refinement of the argument leads to the usual asymptotic development,

$$\Gamma_1(\beta) = 1 + \frac{1}{12\beta} + \cdots. \tag{19}$$

3. **History.** The first proofs of Stirling's formula were given by de Moivre (1730) and Stirling

(1730). Both used what is now called the Euler–MacLaurin formula to approximate $\log 2 + \log 3 + \cdots + \log n$. De Moivre proved the result on the way to the normal approximation for the binomial distribution. His first derivation did not explicitly determine the constant $\sqrt{2\pi}$. In a 1731 addendum, he acknowledged that Stirling was able to determine the constant, using Wallis' formula. To statisticians, the most familiar version of this argument is Robbins (1955); also see Feller (1968). The approach yields upper and lower bounds for $n!$ but does not extend to $\Gamma(\alpha)$.

In essence, we are using a direct form of Laplace's (1774) method to estimate the integral in (12), with quite explicit bounds. For a more general treatment, see de Bruijn (1981, sect. 4.5). For one very similar to ours, see Woodroofe (1975, p. 127). We know of three other approaches to proving Stirling's formula. Modern analysts extend Γ into the complex plane, and have a proof of (1) using the saddlepoint method: see de Bruijn (1981, sect. 6.9). Artin (1964) presents a fascinating discussion of the Γ-function and its properties, as well as a proof of Stirling's formula based on the following theorem: $\Gamma(\alpha)$ *is the only log convex function on* $(0, \infty)$ *satisfying* $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$, *with* $\Gamma(1) = 1$. A third approach using the residue calculus is due to Lindelof; for a modern exposition, see Ahlfors (1979).

We stumbled on our proof while working on finite forms of de Finetti's theorem for exponential families (Diaconis and Freedman, 1980, 1984). As an example, we were thinking of the gamma shape parameter $\alpha > 0$ in the family

$$\frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} \quad \text{for } 0 \leqslant x < \infty. \tag{20}$$

As $\alpha \to \infty$, the density (20) tends to normal, with mean β and variance β. The argument for Stirling's formula was a by-product.

References

1. L. Ahlfors, Complex Analysis, 3rd ed., McGraw-Hill, New York, 1979.
2. E. Artin, The Gamma Function, Holt, Rinehart, Winston, New York, 1964.
3. N. G. de Bruijn, Asymptotic Methods in Analysis, Dover, New York, 1981.
4. A. de Moivre, Miscellanea Analytica de Seriebus et Quadraturis, London, 1730.
5. P. Diaconis and D. Freedman, Finite exchangeable sequences, Ann. Probab., 8 (1980) 745–764.
6. ______, Finite versions of de Finetti's theorem for certain exponential families, Technical report, University of California, Berkeley, 1984.
7. W. Feller, An Introduction to Probability Theory and Its Application, Vol. 1, 3rd ed., Wiley, New York, 1968.
8. P. S. Laplace, Mémoire sur la probabilité des causes par les événements. Mémoires de mathématique et de Physique Presentés à l'Académie Royale des Sciences, par Divers Savants, & Lus dans ses Assemblées, 6, 1774 (Reprinted in Laplace's Oeuvres Complètes. 8, 27–65, English translation by S. Stigler, Technical Report #164, Department of Statistics, University of Chicago, September, 1984.)
9. H. Robbins, A remark on Stirling's formula, this MONTHLY, 62 (1955) 26–29.
10. J. Stirling, Methodus Differentialis, London, 1730.
11. M. Woodroofe, Probability with applications, McGraw-Hill, New York, 1975.

A BOUND FOR THE NUMBER OF MULTIPLICATIVE PARTITIONS

L. E. MATTICS AND F. W. DODD
Department of Mathematics, University of South Alabama, Mobile, AL 36688

Two factorizations of a positive integer are considered to be essentially the same if they differ only in the order of the factors. For example, the four essentially different factorizations of 12 are 12, $6 \cdot 2$, $4 \cdot 3$, and $2 \cdot 2 \cdot 3$. For a positive integer n, let $f(n)$ be the number of essentially different factorizations of n. In a recent note [1] in this MONTHLY, J. F. Hughes and J. O. Shallit proved that $f(n) \leqslant 2n^{\sqrt{2}}$ and conjectured that $f(n) \leqslant n$. We show that this conjecture can be settled in the affirmative in a manner much simpler than they expected.

Let $p_1 = 2, p_2 = 3, \ldots$ be the sequence of primes. We will need the following two lemmas.

LEMMA 1. $\prod_{i=1}^{m}(1 - 1/p_i) \geqslant 1/p_m, m = 1, 2, \ldots$.

LEMMA 2. *If the prime factors of the positive integer $b > 1$ are $p_1, p_2, \ldots, p_m$, then*

$$\sum_{d|b} 1/d < 1 \Big/ \prod_{i=1}^{m}(1 - 1/p_i).$$

Lemma 1 follows from the result $p_{i-1} \leqslant p_i - 1, i \geqslant 2$, and Lemma 2 follows from the result

$$\sum_{d|b} 1/d < \prod_{i=1}^{m}\left(\sum_{k=0}^{\infty}(1/p_i)^k\right).$$

Using these two lemmas we now proceed to settle the conjecture by induction. Since $f(1) = f(2) = f(3) = 1$, the conjecture holds for the first few positive integers. Suppose the conjecture holds for all positive integers less than n. If $n = q_1^{a_1} q_2^{a_2} \cdots q_m^{a_m}, a_1 \geqslant a_2 \geqslant \cdots \geqslant a_m$, is the prime factorization of n and $c = p_1^{a_1} p_2^{a_2} \cdots p_m^{a_m}$, then $f(n) = f(c)$ and $c \leqslant n$. Thus if $c < n$, then $f(n) = f(c) \leqslant c < n$ by the induction hypothesis and we would be done.

There remains the case $n = c$. We further subdivide this case into the subcases $a_m = 1$ and $a_m > 1$. If $m = a_m = 1$, then $n = 2$, a value of n for which we know the conjecture to be true. If $a_m = 1$ and $m \geqslant 2$, then every factorization of n has precisely one factor divisible by p_m so that, if $b = n/p_m$, we have

$$f(n) = \sum_{d|b} f(b/d) \leqslant b \sum_{d|b} 1/d < b \Big/ \prod_{i=1}^{m-1}(1 - 1/p_i) \leqslant b \cdot p_{m-1} < n,$$

by the induction hypothesis and the two lemmas.

Finally, if $a_m > 1$, then set $e = (n/p_m)p_{m+1}$. For any factorization of n, say $n = d_1 d_2 \cdots d_s$, where d_1 is a largest factor divisible by p_m, then $((d_1/p_m)p_{m+1})d_2 d_3 \cdots d_s = e$. Thus essentially different factorizations of n yield in this manner essentially different factorizations of e, so we have $f(n) \leqslant f(e)$. Similar to before, if $b = e/p_{m+1}$, we have

$$f(n) \leqslant f(e) = \sum_{d|b} f(b/d) \leqslant b \sum_{d|b} 1/d < b \Big/ \prod_{i=1}^{m}(1 - 1/p_i) \leqslant b \cdot p_m = n,$$

and the induction is complete.

Reference

1. J. F. Hughes and J. O. Shallit, On the number of multiplicative partitions, this MONTHLY, 90 (1983) 468–471.

GEOMETRIC SERIES AND A PROBABILITY PROBLEM

CURTIS COOPER
Department of Mathematics and Computer Science, Central Missouri State University, Warrensburg, MO 64093

Suppose

$$a_1 + a_2 + a_3 + \cdots$$

is a geometric series with ratio r, $|r| < 1$. We know that the sum of this series is

$$\frac{a_1}{1 - r}.$$

Now if n and $k_1, k_2, \ldots, k_n$ are positive integers, what is $\sum_{i_1 < i_2 < \cdots < i_n} a_{i_1}^{k_1} a_{i_2}^{k_2} \cdots a_{i_n}^{k_n}$? In this

note we will answer this question and then use the result to solve a probability problem.

THEOREM. *Let*

$$a_1 + a_2 + a_3 + \cdots$$

be a geometric series with ratio r, $|r| < 1$. *Let* n *and* $k_1, k_2, \ldots, k_n$ *be positive integers. Then*

$$\sum_{i_1 < i_2 < \cdots < i_n} a_{i_1}^{k_1} a_{i_2}^{k_2} \cdots a_{i_n}^{k_n} = a_1^{k_1+k_2+\cdots+k_n} r^{k_2+2k_3+\cdots+(n-1)k_n} \frac{1}{1 - r^{k_1+\cdots+k_n}} \times \frac{1}{1 - r^{k_2+\cdots+k_n}} \cdots \frac{1}{1 - r^{k_n}}.$$

Proof. By induction on n.

$$a_1^k + a_2^k + a_3^k + \cdots = \frac{a_1^k}{1 - r^k},$$

so the result is true for $n = 1$.

Assume the result is true for $n = m$. Then

$$\begin{aligned}
\sum_{i_1 < i_2 < \cdots < i_{m+1}} a_{i_1}^{k_1} a_{i_2}^{k_2} \cdots a_{i_{m+1}}^{k_{m+1}} &= \sum_{i_1 < i_2 < \cdots < i_m} a_{i_1}^{k_1} a_{i_2}^{k_2} \cdots a_{i_m}^{k_m} \left(a_{i_m+1}^{k_{m+1}} + a_{i_m+2}^{k_{m+1}} + \cdots \right) \\
&= \sum_{i_1 < i_2 < \cdots < i_m} a_{i_1}^{k_1} a_{i_2}^{k_2} \cdots a_{i_m}^{k_m} a_{i_m+1}^{k_{m+1}} (1 + r^{k_{m+1}} + \cdots) \\
&= \sum_{i_1 < i_2 < \cdots < i_m} a_{i_1}^{k_1} a_{i_2}^{k_2} \cdots a_{i_m}^{k_m} (r a_{i_m})^{k_{m+1}} (1 + r^{k_{m+1}} + \cdots) \\
&= \frac{r^{k_{m+1}}}{1 - r^{k_{m+1}}} \sum_{i_1 < i_2 < \cdots < i_m} a_{i_1}^{k_1} a_{i_2}^{k_2} \cdots a_{i_m}^{k_m + k_{m+1}} \\
&= a_1^{k_1+k_2+\cdots+k_{m+1}} r^{k_2+2k_3+\cdots+mk_{m+1}} \frac{1}{1 - r^{k_1+\cdots+k_{m+1}}} \\
&\quad \times \frac{1}{1 - r^{k_2+\cdots+k_{m+1}}} \cdots \frac{1}{1 - r^{k_{m+1}}}.
\end{aligned}$$

Thus, the theorem is true by induction.

We now use this theorem to solve the following problem:

Three players, A, B, and C, take turns throwing a single die, A leads. As soon as a player tosses a one, that player drops out of the game and the remaining players continue rolling the die until everyone has rolled a one. What is the probability that A tosses the first one, B tosses the second one, and C tosses the third one?

To solve this problem, let

$$\frac{1}{6} + \frac{1}{6} \cdot \frac{5}{6} + \frac{1}{6} \cdot \left(\frac{5}{6}\right)^2 + \cdots$$

be the geometric series $a_1 + a_2 + a_3 + \cdots$, where a_i is the probability that player A (,B, or C) throws a one for the first time on the ith toss. The probability that the game ends after A's ith toss, B's jth toss, and C's kth toss is $a_i a_j a_k$. In addition, A will roll the first one, B will roll the second one, and C will roll the third one if and only if $i < j < k$, $i < j = k$, $i = j < k$, or $i = j = k$. Therefore, applying the theorem, the probability that the order of finish is A, B, and C is

$$\sum_{i<j<k} a_i a_j a_k + \sum_{i<j} a_i a_j^2 + \sum_{i<k} a_i^2 a_k + \sum_i a_i^3 = \frac{125}{1001} + \frac{25}{1001} + \frac{5}{91} + \frac{1}{91} = \frac{216}{1001}.$$

THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

Material for this department should be sent to Professor Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.

CONTINUITY OF MEASURABLE FUNCTIONS

A. C. ZAANEN
Department of Mathematics, Leiden University, Leiden, The Netherlands
(Home address: Nassaulaan 15, 2628 GA Delft, The Netherlands)

Let X be a closed interval $[a, b]$ in $\mathbb{R}$ (with $a < b$) and let μ be Lebesgue measure on X. We shall denote by $L_0(X, \mu)$ the set of all complex valued Lebesgue measurable functions on X, with identification of μ-almost equal functions. This identification implies that, strictly speaking, the elements of the set $L_0(X, \mu)$ are not the individual Lebesgue measurable functions, but equivalence classes of such functions. Two functions are "identified", i.e., they belong to the same equivalence class, whenever they differ only on a set of measure zero. In the literature (and in most textbooks) on measure and integration it is customary to make no notational distinction between a measurable function f and the equivalence class to which f belongs. Also one often speaks about a "given measurable function f" when actually it would perhaps be more appropriate to speak about the corresponding equivalence class, which we shall denote by $f^{\sim}$. As long as we deal with statements holding for all $f \in f^{\sim}$ simultaneously, it does no harm to forget about the distinction. As an example, if f_1 and f_2 both belong to $f^{\sim}$, and f_1 is Lebesgue summable, i.e., $\int_X f_1 \, d\mu$ exists as a finite number, then so does $\int_X f_2 \, d\mu$, and the two integrals have the same value. There are other statements, however, which do not hold for all $f \in f^{\sim}$ simultaneously. In textbooks on Fourier analysis, for example, one sometimes finds the assumption that f is a 2π-periodic function which is Lebesgue summable over $[0, 2\pi]$ and has the further property that f is continuous at a given point x_0. What does this mean? It certainly does not mean that all $f \in f^{\sim}$ are continuous at x_0 (this would be nonsense). What comes to mind first is the following definition.

DEFINITION 1. The measurable function f (or rather the equivalence class $f^{\sim}$) on $[a, b]$ is said to be continuous at the point $x_0 \in [a, b]$ if there exists at least one $f_0 \in f^{\sim}$ such that f_0 is continuous at x_0.

Obviously, f_0 is not uniquely determined (because if $x_n \to x_0$ and f_0^1 is defined by $f_0^1(x_n) = f_0(x_n) + n^{-1}$ for all n and $f_0^1(x) = f_0(x)$ for all other x, then $f_0^1 \in f^{\sim}$ and f_0^1 is continuous at x_0). The limit value $f_0(x_0)$ is uniquely determined (i.e., if both f_0 and f_0^1 belong to $f^{\sim}$ and both are continuous at x_0, then $f_0(x_0) = f_0^1(x_0)$). There is still another possibility for defining continuity of $f^{\sim}$ at x_0. Some readers may prefer the (ε, δ)-definition which we present now.

DEFINITION 2. The measurable function f (or rather $f^{\sim}$) on $[a, b]$ is said to be continuous at x_0 if there is a number α_0 having the property that for every $\varepsilon > 0$ there exists $\delta > 0$ such that if $f \in f^{\sim}$, then $|\alpha_0 - f(x)| \leqslant \varepsilon$ for μ-almost all $x \in [a, b]$ satisfying $|x - x_0| \leqslant \delta$ (the exceptional set of measure zero depending on the choice of f).

Are the two definitions equivalent? It is evident that if $f^{\sim}$ satisfies the first definition, then $f^{\sim}$ also satisfies the second definition. Conversely, assume that $f^{\sim}$ satisfies Definition 2. Take a sequence $\varepsilon_n \downarrow 0$ and the corresponding sequence δ_n. Then choose an arbitrary $f \in f^{\sim}$ and denote by E_n the subset of $(x : |x - x_0| \leqslant \delta_n)$ where $|\alpha_0 - f(x)| \leqslant \varepsilon_n$ is not satisfied. The set E_n is of

measure zero, and hence $E = \bigcup_1^\infty E_n$ is of measure zero. Define $f_1 \in f^\sim$ by $f_1(x) = \alpha_0$ for all $x \in E$ and $f_1(x) = f(x)$ elsewhere. It is easy to see that f_1 is continuous at x_0, which shows that $f^\sim$ satisfies Definition 1. It follows at the same time that the number α_0 in Definition 2 is uniquely determined.

Assume now that $f^\sim$ is continuous at all $x \in [a, b]$ according to the (equivalent) definitions above. This means that to each $x \in [a, b]$ there is assigned a number α_x and a function $f_x \in f^\sim$ such that $f_x(x) = \alpha_x$ and f_x is continuous at the point x. Of course, all these f_x are (in general) mutually different functions, the more so because no f_x is uniquely determined. The question arises now whether or not there exists a function $f_0 \in f^\sim$ which is continuous at all $x \in [a, b]$ simultaneously, because if not, this would contradict the usual meaning of the statement that a "function" is continuous on $[a, b]$. An obvious candidate for a continuous $f_0 \in f^\sim$ is the function α, defined by $\alpha(x) = \alpha_x(x)$ for all $x \in [a, b]$. To show that this is a good conjecture, we have to prove firstly that α is continuous on $[a, b]$ and secondly that α belongs to $f^\sim$.

(i) Assume that α fails to be continuous at some point x_0. Then there exist a number $\varepsilon > 0$ and a sequence $x_n \to x_0$ such that $|\alpha(x_n) - \alpha(x_0)| > \varepsilon$ for all n, i.e.,

$$|\alpha_{x_n}(x_n) - \alpha_{x_0}(x_0)| > \varepsilon \quad \text{for all } n. \tag{1}$$

Choose an arbitrary $f \in f^\sim$. There exists a number $\delta > 0$ such that for μ-almost every y satisfying $|y - x_0| \leqslant \delta$ we have

$$|f(y) - \alpha_{x_0}(x_0)| < \frac{1}{2}\varepsilon. \tag{2}$$

We may assume that $|x_n - x_0| < \frac{1}{2}\delta$ for all n. Corresponding to x_n and ε there exists a number $\delta_n > 0$ such that for μ-almost every y satisfying $|y - x_n| \leqslant \delta_n$ we have

$$|f(y) - \alpha_{x_n}(x_n)| < \frac{1}{2}\varepsilon, \tag{3}$$

where we may assume that $\delta_n < \frac{1}{2}\delta$ (so all y in (3) satisfy $|y - x_0| < \delta$). From (2) and (3) it follows now that the left-hand side in (1) is less than ε. Contradiction. Hence, the function α is continuous on $[a, b]$.

(ii) Choose $\varepsilon > 0$ and an arbitrary $f \in f^\sim$. For each $x \in [a, b]$ there exists an open interval A_x with center x such that both $|\alpha(y) - \alpha(x)| < \frac{1}{2}\varepsilon$ and $|f(y) - \alpha(x)| < \frac{1}{2}\varepsilon$ for μ-almost every y in $A_x \cap [a, b]$. Hence

$$|f(y) - \alpha(y)| < \varepsilon \quad \text{for } \mu\text{-almost all } y \in A_x \cap [a, b].$$

Since the open covering $\{A_x\}$ of $[a, b]$ has a finite subcovering, we find that $|f(y) - \alpha(y)| < \varepsilon$ for μ-almost every $y \in [a, b]$. Now let $\varepsilon_n \downarrow 0$. Then $|f(y) - \alpha(y)| < \varepsilon_n$ except on a set E_n of measure zero. Hence $|f(y) - \alpha(y)| < \varepsilon_n$ for all ε_n simultaneously except on the set $E = \bigcup_1^\infty E_n$ of measure zero. It follows that f and α are equal except on a set of measure zero, i.e., $\alpha \in f^\sim$.

We briefly look at the analogous questions concerning differentiability. One way to say that $f^\sim$ is differentiable at x_0 is that there exists at least one $f_0 \in f^\sim$ such that f_0 is differentiable at x_0 (and hence f_0 is continuous at x_0). Equivalently, there exist numbers α_0 and α_0' with the property that for any $\varepsilon > 0$ there exists $\delta > 0$ such that if $f \in f^\sim$, then

$$\left| \frac{f(y) - \alpha_0}{y - x_0} - \alpha_0' \right| < \varepsilon$$

for μ-almost every y satisfying $0 < |y - x_0| \leqslant \delta$. We leave it to the reader to prove the equivalence.

Assume now that $f^\sim$ is differentiable at each $x \in [a, b]$, i.e., for each $x \in [a, b]$ there exists a

function $f_x \in f^{\sim}$ such that f_x is differentiable at x. The question which arises is again whether there is a function in $f^{\sim}$ which is differentiable on the whole interval $[a, b]$, and hence continuous on $[a, b]$. If so, the only candidate is the function α, defined by $\alpha(x) = \alpha_x(x)$, which we have met already above. For the proof, choose an arbitrary $x \in (a, b)$; the proof for $x = a$ or $x = b$ is similar. There exists $f_x \in f^{\sim}$ such that f_x is differentiable at x. Hence, given $\varepsilon > 0$, there exists $\delta > 0$ such that if $0 < |y - x| < \delta$, then

$$\left| \frac{f_x(y) - \alpha(x)}{y - x} - \alpha'_x \right| < \varepsilon,$$

where we have used that $f_x(x) = \alpha(x)$. Since $f_x(y) = \alpha(y)$ for μ-almost all $y \in [a, b]$, we get

$$\left| \frac{\alpha(y) - \alpha(x)}{y - x} - \alpha'_x \right| < \varepsilon \tag{4}$$

for μ-almost every y satisfying $0 < |y - x| < \delta$. Let us consider first the interval $A = (y{:}0 < y - x < \delta)$. On the interval A the left-hand side of (4), call it $\beta(y)$, is a continuous function satisfying $0 \leqslant \beta(y) < \varepsilon$, for almost all $y \in A$. In view of the continuity it follows that $\beta(y) \leqslant \varepsilon$ for all $y \in A$. A similar statement holds for $(y{:} -\delta < y - x < 0)$. Hence $\beta(y) \leqslant \varepsilon$ for $0 < |y - x| < \delta$. This shows that α is differentiable at the point x.

These remarks may serve to show that so far as questions of continuity or differentiability of Lebesgue measurable functions are concerned, no harm arises if we continue to handle equivalence classes as if they were functions, thereby writing f again instead of $f^{\sim}$.

ON POSING PROBLEMS IN MATHEMATICS

STEVEN H. WEINTRAUB
Department of Mathematics, Louisiana State University, Baton Rouge, LA 70803-4918

In this note I would like to report on a technique I have used to advantage in the teaching of graduate and upper-level undergraduate courses. In such a course one often wants to pose interesting and challenging problems, e.g., proving useful results related to theorems presented in class, or working out non-trivial examples of a general theory, but is faced with the difficulty that many of the most interesting problems are simply *too hard* for the students to answer. My technique is one that to an extent circumvents this obstacle, and allows one to pose otherwise inaccessible problems.

The technique is simply the following: At any time (before the problem is due) a student can ask me a "yes-or-no" question about it. Of course, such a question might be just "Is what I've done so far correct?" (Note, though, that "What have I done wrong?" is not a legitimate question.) However, the power of this technique comes in some of the other questions that can be asked. For example, a student may ask "Is such-and-such true?" (though not, of course, "How do I prove such-and-such?"). As we all know, it's a lot easier to prove something when you know it's true, so a "yes" answer here would materially help the student (and a "no" answer might avert an extended sojourn in a blind alley). Also, a sequence of such questions would enable the student to sketch the outlines of a difficult argument, and then go back and work on the proofs of the individual steps. (This sort of decomposition of a problem into subproblems is itself an important method for the student to learn.)

As with any technique, this one must be applied with a certain amount of judgement (e.g., sometimes I will reply to a "yes-or-no" question by saying "Think about it some more"), but I have indeed found it to be useful.

PROBLEMS AND SOLUTIONS

EDITED BY G. L. ALEXANDERSON, H. M. W. EDGAR (ELEMENTARY PROBLEMS),
D. H. MUGLER, AND KENNETH B. STOLARSKY (ADVANCED PROBLEMS)

*Send all **proposed** problems, typed and in duplicate if possible, to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the address given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by June 30, 1986. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgement).

E 3129. *Proposed by R. Kannan, Carnegie-Mellon University; D. J. Kleitman, Massachusetts Institute of Technology; and J. C. Lagarias, AT & T Bell Laboratories.*

Let $f(n)$ denote the maximum number of disjoint rectangles that the unit square in $\mathbb{R}^2$ can be partitioned into such that any horizontal line in $\mathbb{R}^2$ intersects the interior of at most n rectangles and any vertical line in $\mathbb{R}^2$ intersects the interior of at most n rectangles.

(a) Show that

$$3 \cdot 2^{n-1} - 2 \leqslant f(n) < 3^{n+3}.$$

(b)* Prove or disprove that $f(n) = 3 \cdot 2^{n-1} - 2$, i.e., that $f(n)$ is determined by $f(1) = 1$ and the recurrence

$$f(n) = 2f(n-1) + 2.$$

E 3130. *Proposed by Mark Kantrowitz (student), Brookline, MA.*

Let $\{H_n\}$ be a generalized Fibonacci sequence, i.e., H_1 and H_2 are arbitrary integers and for $n > 2, H_n = H_{n-1} + H_{n-2}$.

(a) Find T, in terms of H_1 and H_2, such that

$$H_{2n}H_{2n+2} + T, \quad H_{2n}H_{2n+4} + T, \quad H_{2n-1}H_{2n+1} - T, \quad \text{and} \quad H_{2n-1}H_{2n+3} - T$$

are all perfect squares.

(b) Prove that T is unique.

E 3131. *Proposed by William N. Anderson, Jr., Fairleigh Dickinson University, and George E. Trapp, West Virginia University.*

Let A be a partitioned real matrix with submatrices $A_{ij}, i, j = 1, \dots, k$. Let B be the $k \times k$ matrix with elements b_{ij} given as follows: b_{ij} is the algebraic sum of all of the elements in the A_{ij} submatrix. Show that if A is symmetric positive definite, then B is symmetric positive definite.

E 3132. *Proposed by Robert E. Shafer, Berkeley, CA.*

For integers $n \geqslant 2$ and real $s > 0$, show that

$$\left(\prod_{i=0}^{n}(s+i)\right)\left(\sum_{j=0}^{n}\frac{1}{s+j}\right) < (n+1)\prod_{k=1}^{n}\left(s+k-\frac{1}{2}\right).$$

E 3133. *Proposed by Peter D. Zvengrowski, The University of Calgary.*

Let $\Delta^n \subset \mathbb{R}^n$ be a regular n-simplex with center at 0. Let $a_i = (a_{i1}, a_{i2}, \dots, a_{in})$ be its vertices, $i = 0, 1, \dots, n$. Show that for all j and k, $\sum_{i=0}^{n} a_{ij}a_{ik} = c\delta_{jk}$ for some constant c, where δ_{jk} is the Kronecker delta.

E 3134. *Proposed by Jordi Dou, Barcelona, Spain.*

Provide a Euclidean construction of a triangle ABC given the median m_A, the bisector t_A and the angle A.

SOLUTIONS OF ELEMENTARY PROBLEMS

Simplices, Circumspheres, and Centroids

E 2987 [1983, 133; 1985, 669] (**Addendum**).
The following was inadvertently omitted from the end of the solution of E 2987:

The spheres for different K are clearly nonintersecting and if $K = n + 1$ we obviously get the sphere with OG as diameter. Further, f^2 is a quadratic function of $x = \dfrac{n+1}{K+n+1}$ which has a minimum at $x = \dfrac{R^2}{OG^2} > 1$. If $K > 0$ and increasing, then $x < 1$ and decreasing so f is increasing, which proves the result.

Also solved by A. Bager (Denmark), J. M. Becker (France), I. J. Schoenberg, and the proposer.

Isogonal and Isotomic Conjugates

E 2990 [1983, 212]. *Proposed by H. Eves, University of Maine, and C. Kimberling, University of Evansville.*

Let ABC be a triangle and L a line in the plane of ABC not passing through A, B, or C.

(i) Prove that the isogonal conjugate of L is an ellipse, parabola or hyperbola according as L meets the circumcircle of ABC in zero, one or two points.
(ii) Prove that the isotomic conjugate of L is an ellipse, parabola, or hyperbola according as L meets E in zero, one, or two points, where E is the ellipse through A, B, C having the centroid of triangle ABC as center.

Composite solution by O. P. Lossers, Eindhoven University of Technology, The Netherlands, and Peter Yff, American University of Beirut, Lebanon.

(i) If P has trilinear coordinates (x, y, z) with ABC as triangle of reference, then P^*, the isogonal conjugate of P, has coordinates proportional to $(1/x, 1/y, 1/z)$. Thus, if L has equation $px + qy + rz = 0$, then its isogonal conjugate is the conic L^* with equation $pyz + qzx + rxy = 0$. The isogonal conjugate S^* of the circumcircle S is the line at infinity. Since $L^* \cap S^*$ and $L \cap S$ have the same number of points, L^* is an ellipse, parabola, or hyperbola depending on whether $L \cap S$ contains zero, one, or two points, as we wanted to prove.

(ii) Under an affine transformation sending ABC to an equilateral triangle $A'B'C'$, L is mapped to a line L' and E is mapped to the circumcircle of $A'B'C'$. But the isotomic conjugate of L' with respect to $A'B'C'$ is also its isogonal conjugate. Consequently, using the fact that isotomic conjugacy is an affine invariant, we obtain the required result from part (i).

Part (ii) can also be solved exactly as part (i), using barycentric coordinates in place of trilinear coordinates. Indeed, it was observed by the proposers and Peter Yff that the results of (i) and (ii) can be generalized by introducing any system of homogeneous coordinates and considering the image of L under the transformation sending (x, y, z) to $(p/x, q/y, r/z)$, where p, q, r are constants.

L. M. Kelly noted that part (i) appears in the classical literature, e.g., in the works of John Casey.

Also solved by J. Dou (Spain), R. H. Eddy (Canada), L. M. Kelly, L. Kuipers (Switzerland), and the proposers. A partial solution was given by H. E. Fettis.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by June 30, 1986. The solver's full post-office address should be on each sheet.

6509. *Proposed by Boo Rim Choe, University of Wisconsin at Madison.*

Let $D = \{r_i\}_1^\infty$ be a countable dense subset of $\mathbb{R}$ and $\{a_n\}_1^\infty$ a sequence of positive real numbers. If $\lim_{n\to\infty} a_n = 0$, show that a sequence $f_n\colon \mathbb{R} \to \mathbb{R}$ of positive continuous functions can be constructed such that

$$\lim_{n\to\infty} f_n(x) = \begin{cases} +\infty & \text{if } x \notin D, \\ a_m^{-1} & \text{if } x = r_m \in D. \end{cases}$$

Also, what can be said if a_n does not converge to zero?

6510. *Proposed by S. V. Kanetkar, Old Dominion University, Norfolk, VA.*

Let G be a finite group with nontrivial cyclic Sylow 2-groups. Prove that the product of all the elements never equals the identity. (Generalization of Wilson's Theorem.)

SOLUTIONS OF ADVANCED PROBLEMS

Lifting an Endomorphism to an Automorphism

6471 [1984, 519]. *Proposed by Justin Peters, Iowa State University.*

Let R be a ring and α an injective endomorphism of R; i.e., $\alpha\colon R \to R$ is a 1-1 ring homomorphism, not necessarily onto. Show that there is a ring S containing R as a subring and an automorphism β of S such that $\beta(x) = \alpha(x)$ for all $x \in R$.

Solution by David A. Sibley, Pennsylvania State University, University Park. Let $T = R^{\mathbb{N}}$, which is a ring under co-ordinatewise operations. Define $f\colon R \to T$ via

$$f(r) = \big(r, \alpha(r), \alpha^2(r), \dots\big) \quad \text{for all } r \in R.$$

Clearly, f is a monomorphism, so $R \cong f(R)$. Let I be the ideal of T consisting of all sequences which are 0 except at a finite number of positions and set $S = T/I$. Since $f(R) \cap I$ is zero, we have

$$\big(f(R) + I\big)/I \cong f(R) \cong R.$$

Let l be the left shift operator on T. Since I is l-invariant, l induces an endomorphism $\bar{l}$ of S via $\bar{l}(t + I) = l(t) + I$ for all $t \in T$. It is easy to verify that $\bar{l}$ is an automorphism of S. Furthermore, as $l(f(r)) = f(\alpha(r))$ we have

$$\bar{l}\big(f(r) + I\big) = f\big(\alpha(r)\big) + I \quad \text{for all } r \in R.$$

That is, S has a subring $(f(R) + I)/I$ isomorphic to R and an automorphism $\bar{l}$ whose restriction to this subring is the same as α, as required.

Comment: J. Clark (University of Otago, New Zealand) points out that D. A. Jordan (Bijective extensions of injective ring endomorphisms, *J. London Math. Soc.* (2) 25 (1982) no. 3, 435–448) constructs an overring S with the desired properties. In fact, in the overring S thus constructed, β is an inner automorphism of S (i.e., there is an invertible $s \in S$ for which $\beta(x) = sxs^{-1}$ for $x \in S$). This article also discusses how certain properties pass from R to S.

Also solved by Erich Badertscher (Switzerland), F. Rudolf Beyl, Alberto Facchini (Italy), A. A. Jagers (The Netherlands), O. P. Lossers (The Netherlands), Victor Pambuccian (Romania), Brian Peterson, Daniel Zelinsky, and the proposer.

Operators on Hilbert Space

6472 [1984, 519]. *Proposed by Robin Harte, University College, Cork, Ireland.*

Prove that if $T\colon H \to H$ is a bounded linear operator on Hilbert space for which $T^2 = 0$, then the following two conditions are equivalent:
(a) $I = UT + TV$ for some bounded linear U and V on H;
(b) $T + T^*$ is invertible.

Solution by J. Everett and K. Kearnes, University of California, Berkeley.
(i) Suppose (a) holds. First, $T^2 = 0$ implies $\operatorname{im} T \subseteq \ker T$. Now, if $x \in \ker T$, then

$$x = UTx + TVx = TVx \in \operatorname{im} T,$$

so $\operatorname{im} T = \ker T$ (likewise, $\operatorname{im} T^* = \ker T^*$). This, together with the boundedness of T, implies $\operatorname{im} T$ is closed. We conclude that

$$H = \operatorname{im} T \oplus (\operatorname{im} T)^{\perp} = \operatorname{im} T \oplus \ker T^* = \operatorname{im} T \oplus \operatorname{im} T^*.$$

To show $T + T^*$ is 1-1, we choose $x \in \ker(T + T^*)$. Then

$$Tx = -T^*x \quad \text{so} \quad Tx, T^*x \in \operatorname{im} T \cap \operatorname{im} T^* = \{0\},$$

and thus $x \in \ker T \cap \ker T^* = \{0\}$. To show $T + T^*$ is onto, it suffices to show $\operatorname{im} T \subseteq \operatorname{im}(T + T^*)$ since then, by symmetry, $\operatorname{im} T^*$ (and hence H) is also contained in $\operatorname{im}(T + T^*)$. If $r = Tx$ where $x = Ty + T^*z$, then

$$r = T(Ty + T^*z) = TT^*z = (T + T^*)T^*z \in \operatorname{im}(T + T^*).$$

Thus, $T + T^*$ is invertible.

(ii) Suppose (b) holds. Let $W = (T + T^*)^2 = T^*T + TT^*$. Then W is invertible and both W and W^{-1} commute with T. Thus,

$$I = W^{-1}W = W^{-1}(T^*T + TT^*) = (W^{-1}T^*)T + T(W^{-1}T^*),$$

and hence $I = UT + TV$ with $U = V = W^{-1}T^*$, which is bounded since T is.

Also solved by Erich Badertscher (Switzerland), F. Rudolf Beyl, Rafael Obaya Garcia (Spain), Scott H. Hochwald, O. P. Lossers (The Netherlands), Pei Yuan Wu (Republic of China), and the proposer.

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER

Astronomical Cuneiform Texts, Vols. I, II, and III. By O. Neugebauer. Sources in the History of Mathematical and Physical Sciences, Number 5. Springer-Verlag, New York, 1983. Vol. I, xvii + 278 pp.; Vol. II, x + 233 pp.; Vol. III, 255 pp., $88.00.

Astronomy and History, Selected Essays. By O. Neugebauer. Springer-Verlag, New York, 1983. xii + 538 pp., $19.80.

N. M. SWERDLOW
Department of Astronomy and Astrophysics, The University of Chicago, Chicago, IL 60637

Astronomy, after mathematics, is the oldest of the exact sciences, although it is not nearly so old. There are cuneiform texts from early in the second millennium B.C. that already show the full development of Babylonian mathematics, but astronomy of comparable sophistication does not appear until some fifteen centuries later. Like mathematics, Babylonian astronomy was intended for practical purposes. The mathematical texts mostly contain problems dealing with digging ditches, building walls, compound interest, inheritance, and other matters pertinent to life in Mesopotamia. The purpose of the astronomical texts appears, to us at least, somewhat stranger. From an early period certain characteristic *phenomena* of planets, such as heliacal risings (the first rising of a star after having been invisible because of its nearness to the sun) and settings, stations and oppositions, the first and last visibilities and oppositions of the moon, and eclipses of the moon and sun were regarded as omens by which the gods revealed the future course of events and whether particular times were auspicious or inauspicious. One could of course observe such phenomena directly, and this was indeed done at least as early as the second millennium. But since the purpose of observing and recording omens was to take appropriate action, to obtain the advantages of the favorable and avoid the consequences of the unfavorable, it would be useful to learn to predict the time of the phenomena in advance and then plan accordingly. This was the "practical" motivation of what turned out to be the most sophisticated and comprehensive mathematical astronomy of antiquity before the time of Ptolemy.

O. Neugebauer's *Astronomical Cuneiform Texts* is an edition with full technical analysis of all the texts known—about three hundred, mostly fragmentary—at the time of its original publication in 1955. Unavailable for some years, it has now been reprinted by Springer with the addition of a bibliography of more recent publications. It is, quite simply, one of the monuments of the history of science, an accomplishment of enormous industry and considerable genius.

The pioneering work in the decipherment of Babylonian astronomy had been done in the late nineteenth and early twentieth centuries by J. N. Strassmaier, J. Epping, and F. X. Kugler (all Jesuits). Among the texts they considered, many consisted of nothing but columns of numbers following recurring sequences between fixed limits, and these had earlier been classified as

economic texts, i.e., accounts and such. They were correctly identified as astronomical by Strassmaier, and rules for computing many of the columns were worked out by Epping and later Kugler who showed, among much else, that they were designed to predict the very phenomena regarded as omens. In the 1930's, Neugebauer found that by the application of linear diophantine equations, it was possible to form sequences connecting and dating fragments of texts over a period of a few hundred years, making up more-or-less continuous records, called "ephemerides", of each of the phenomena. This, along with the decipherment of many more previously unknown or unintelligible columns, brought with it a deeper understanding of the relation of the hundreds of fragments and made it possible to reconstruct missing sections and produce a systematic edition of the texts classified according to astronomical significance, method of computation, and numerical parameters. The result, some twenty years later, was *ACT*, as it is usually called, which remains to this day and for the foreseeable future the foundation of our knowledge of this remarkable creation of a long vanished civilization.

Babylonian mathematical astronomy appears to have been developed quite rapidly during the fourth and third centuries B.C. by a handful of scribes that must be ranked among the most original of all scientists (a term I use deliberately). Indeed, from the thoroughness and complexity of their work, it is obvious that their interests had gone far beyond the initial object of predicting ominous phenomena to the purely scientific and highly abstract analysis of apparently irregular phenomena into combinations of periodic functions that could be represented numerically. Their fundamental insight was that the phenomena, such as heliacal risings, which varied greatly in synodic arc and time separating successive occurrences, were not random, but followed patterns that would eventually repeat. The periods of these repetitions used in the ephemerides are fairly long, longer than the time for which the Babylonians had useful observations at their disposal, but these long periods are themselves formed from the addition of shorter, although less accurate, periods of at most a couple of generations, through which the errors of the shorter periods are presumed to cancel. For example, ephemerides for Jupiter are usually computed for a cycle of 427 years containing 391 occurrences of each phenomenon and 36 rotations of the planet through the zodiac—note that $427 = 391 + 36$, a necessary relation for superior planets—and this long cycle is in turn based upon shorter cycles, namely,

$$a.\quad 12 = 11 + 1 \quad (+5^\circ), \qquad b.\quad 71 = 65 + 6 \quad (-6^\circ),$$

which are combined as $6a + 5b$ to form the long period. Thus, all that is required to arrive at these periods is the counting of synodic phenomena, as heliacal risings, and a rough estimate of the excess or deficit of the period of some number of phenomena from an integral number of years. The observations need be neither precise nor extensive; a record of just the dates of some of the phenomena spread over about a century is sufficient, and the scribes seemed to have had access to records of three or four centuries of such observations. Their ingenuity was in how they used them, for a basic principle of Babylonian astronomy, as also of later Greek astronomy, was to devise methods of getting the best possible results out of carefully selected observations that did not in themselves have to be particularly accurate.

Within each of these long periods the phenomena are separated by irregular intervals of synodic arc and time, but even a rough comparison of the intervals and zodiacal longitudes of the phenomena is enough to show that the intervals are a function of longitude, that is, in certain parts of the zodiac the phenomena are closely spaced and in others widely spaced, the differences amounting to some number of days. Again, rather crude observations, say, to the nearest day, suffice to establish the limits of the inequality. Two methods used to represent this inequality are shown in Fig. 1, which is drawn for the synodic phenomena of Jupiter. The left-hand scale $\Delta\lambda$ in the figure shows the length of the synodic arc in degrees and the right-hand scale Δt shows the excess of the synodic time over 12 mean synodic months in units of $1/30$ of a mean synodic month. These are conventionally called "tithis" (τ), a term borrowed from Indian astronomy in which this Babylonian unit continued in use. Longitude λ in the bottom scale is shown in zodiacal signs of 30°.

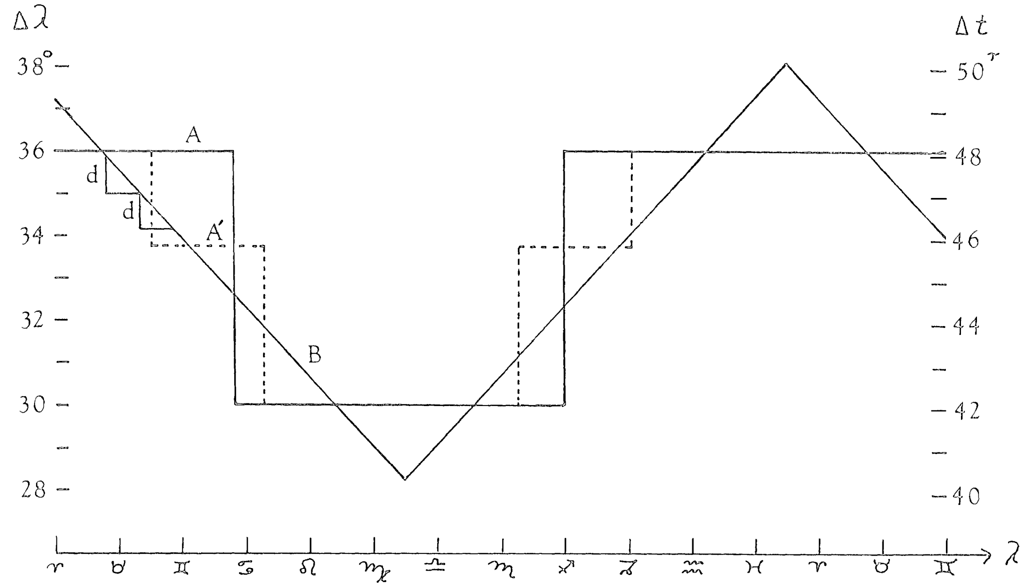


FIG. 1

In one method of representing the inequality, known as System A, the zodiac is divided into from two to as many as six zones, in each of which the synodic arc and time between phenomena is constant. This is called a "step function," and the figure shows two versions for Jupiter, A with arcs of 30° and 36° and A′ with the addition of short transitional arcs of 33; 45°. Corresponding to these are times of 42; 5, 10$^{\tau}$, 48; 5, 10$^{\tau}$, and 45; 50, 10$^{\tau}$, all following from the relation $\Delta t = \Delta\lambda + 12; 5, 10^{\tau}$. In the second method, called System B, the difference d of successive synodic arcs and times is constant, for Jupiter 1; 48° and 1; 48$^{\tau}$, and the variation is contained between specified limits of 28; 15, 30° and 38; 2°, and of 40; 20, 45$^{\tau}$ and 50; 7, 15$^{\tau}$, for arc and time, respectively, the relation here being $\Delta t = \Delta\lambda + 12; 5, 15^{\tau}$. The pattern of System B is called a "linear zigzag function," the reason for which should be evident from the figure. In all cases the accuracy of the parameters, to two and sometimes more fractional places, results from arithmetical considerations in their derivation and does not reflect any such accuracy in observations. In fact, all that was ultimately considered significant was the date of a phenomenon and the zodiacal sign in which it occurred.

The ephemerides, computed by either System A or System B, contain columns of dates and longitudes of each of the synodic phenomena, and sometimes also columns of the intervals of time and arc that form their basis. The computation is quite simple, nothing more than addition is required, and special rules show what to do at the jumps of the step function and limits of the zigzag. They can be continued in principle for one full cycle, after which all intervals of time and arc will repeat, as also will the longitudes, but, due to intercalations in the calendar, not the dates. Surviving texts extend from the fifth century B.C. to the first century A.D., with the greatest concentration in the last two centuries B.C. Now it was just at this time that mathematical astronomy spread from Mesopotamia, eastward to India and westward to the Mediterranean, although in both places it was transformed so completely that few vestiges of Babylonian procedures remained. Nevertheless, some of the numerical parameters continued in use in Greek and Indian astronomy, and simple linear zigzag schemes for determining the variation of the length of daylight survived long into the middle ages.

But very important changes did take place. As is well known, the Greeks approached astronomy, as mathematics, geometrically, asking what model, assumed to be composed of circular motions, could represent the apparent motions of the sun, moon, and planets, and not just the ominous phenomena that were of primary concern to the Babylonians. Instead, the Hellenistic

Greeks developed, apparently along with their astronomy, the fundamentally different procedures of astrology, which is concerned with the relations and locations of the planets in the zodiac, that is, of points on a circle, a subject that appears to have been only of secondary interest to the Babylonians. And while omens occur only at disparate times—most days are not ominous—the configurations of the planets in the zodiac can be considered at any time, offering new and unlimited possibilities for divination, for the "practical" application of astronomy. This may well account for the survival of Greek, and the disappearance of Babylonian, mathematical astronomy. But there is also a significant scientific consequence, for the very object of mathematical astronomy changed from the Babylonian: at what *time* will a given phenomenon occur? to the Greek: at a given time, what is the *location* of a planet? Note that a coordinate system is only secondary, at most a means to an end, in the first case, but is the essence of the second. And it is the second problem that has remained the object of celestial mechanics to the present day.

The earlier parts of this story, for antiquity, are set out in detail in Neugebauer's comprehensive *A History of Ancient Mathematical Astronomy* (*HAMA*), published in three volumes by Springer in 1975. Probably the most well-known book on ancient mathematics and astronomy, and certainly the best introduction to the subject, is Neugebauer's *The Exact Sciences in Antiquity*, originally published in 1951 and currently available in a second edition from Dover. Springer has now brought out a selection of Neugebauer's papers, called *Astronomy and History*, that should be of interest to anyone who has read *The Exact Sciences*. The collection covers not only Egyptian, Babylonian, and Greek astronomy, but also geography, astrology, and later Indian, medieval European, Renaissance, and Ethiopic astronomy and chronology. The volume contains some classic papers that are all but unobtainable today, and is highly recommended. Three of these papers written in the early 1940's, survey, almost non-technically, the entire range of ancient astronomy. Another long paper is an exposition of the very sophisticated astronomy in Maimonides's treatment of the Jewish calendar. Among others that may be noted is one on the development of Greek planetary theory in India, and one on Copernicus containing a deeper understanding of his work than can be found in nearly all of the literature on this rather archaic astronomer, who was far closer to Ptolemy than to Kepler. Finally, some recent papers concerned with Ethiopic astronomy and chronology show an unexpected survival of the early Alexandrian ecclesiastical calendar, predating the ecclesiastical calendar used throughout the middle ages, and show how simple is the astronomical basis of such calendars. There are forty-three papers in all, written over a period of some forty years, and the collection can serve as a good introduction to *HAMA* and *ACT*. The subjects are diverse and, in some cases, quite specific, for it is only through a total command of detail that the full richness of ancient science can be brought to life, as Professor Neugebauer has done to a greater degree than any scholar of our time.

162.

MISCELLANEA

Name This Book

Which book by the author in parentheses is suggested by the given mathematical phrase or symbol?

1. The Fundamental Theorem of Algebra (A. Haley)
2. The Bernoullis (I. Turgenev)
3. Matrix Multiplication (C. Dickens)
4. $\mathrm{Re}(z) > 0$ (T. Wolfe)
5. $\lim_{x\to-\infty} f(x)$ (C. Kingsley)
6. $\lim_{x\to+\infty} f(x)$ (T. Hardy)
7. $\int_0^\infty f(x)\,dx$ (J. Jones)
8. Line Integration (J. Conrad)

The answers appear on p. 144.

Wells Johnson
Bowdoin College
Brunswick, ME 04011

approaches to the same problem. In recent years it has become evident that completely integrable nonlinear ordinary and partial differential equations, which describe important physical processes, admit very detailed analyses and have solutions with remarkable properties. Simple tests now being developed for determining whether a system is completely integrable can be traced back through the work of Painlevé to works of Fuchs and to Kovalevskaia's local study of the complexified problem of rotation. There still remain, however, deep questions on the relationship of her approach to integrability theory. Cf. **[8]** and the references cited therein. (It is interesting to note that Emmy Noether, Kovalevskaia's successor as the commonly acknowledged greatest female mathematician, made major contributions to the problem of integrability in her 1918 study of invariant variational problems.)

Koblitz's book gives a very readable account of Kovalevskaia's life. It was a life filled with personal drama. The author has uncovered much new material about Kovalevskaia. But the author's admiration for her subject and her lack of expertise in mathematics have led her to color many episodes of Kovalevskaia's mathematical life and assessments of her work in ways more favorable to Kovalevskaia and less favorable to her ostensible opponents than a sober evaluation of the evidence and its context warrants. Kovalevskaia does not require such assistance. She was the author of influential papers and was respected by the mathematical community for mathematical abilities beyond those reflected in her research. She was recognized for her literary efforts. But perhaps her greatest legacy is that today there are many female mathematicians whose *mathematical* accomplishments have surpassed hers.

References

1. E. T. Bell, Men of Mathematics, Simon & Schuster, 1937.

2. Y. V. Golubev, Lectures on Integration of the Equations of Motion of a Rigid Body about a Fixed Point, U.S. Dept. Comm., 1960. (English transl. of Russian edition of 1953.)

3. P. J. Holmes and J. E. Marsden, Horseshoes and Arnold diffusion for Hamiltonian systems on Lie groups, Indiana Univ. Math. J., 32 (1983) 273–309.

4. F. Klein, Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert, Heft I, Springer, 1927.

5. F. Klein and A. Sommerfeld, Über die Theorie des Kreisels, Teubner, 1897–1910.

6. M. Kline, Mathematical Thought from Ancient to Modern Times, Oxford Univ. Press, 1972.

7. E. E. Kramer, Sonya Kovalevsky, Dictionary of Scientific Biography, vol. VII, Scribner's, 1973, pp. 447–480.

8. M. Kruskal, Generalized Painlevé tests for integrability of differential equations, in preparation.

9. E. Leimanis, The General Problem of the Motion of Coupled Rigid Bodies about a Fixed Point, Springer-Verlag, 1965.

10. G. Mittag-Leffler, Sophie Kovalevsky, notice biographique, Acta Math., 16 (1983) 385–392.

11. G. Mittag-Leffler, Weierstrass et Sonja Kowalewsky, Acta Math., 39 (1923) 133–198.

ANSWERS TO "NAME THIS BOOK" ON PAGE 138

1. Roots.
2. Fathers and Sons.
3. Hard Times.
4. The Right Stuff.
5. Westward ho!
6. Far from the Madding Crowd.
7. From Here to Eternity.
8. The End of the Tether.

A Convergence of Lives—Sophia Kovalevskaia: Scientist, Writer, Revolutionary. By Ann Hibner Koblitz. Birkhäuser, Boston, 1983. xx + 305 pp.

STUART S. ANTMAN
Department of Mathematics, University of Maryland, College Park, MD 20742

Sonya Kovalevsky was the greatest woman mathematician prior to the twentieth century.

E. E. Kramer [7]

Kowalewsky is one of the few women mathematicians of distinction.

M. Kline [**6**, p. 702]

At the time of her death, Kovalevskaia was indeed considered the equal of anyone of her generation. This included Poincaré, Picard, and Mittag-Leffler.

Ann Hibner Koblitz (p. 250)

Sofia V. Kovalevskaia*, the object of these effusions, was born in Moscow in 1850 to a wealthy landowning family. She grew up on their provincial estate, ineffectively shielded from the intellectual and political ferment then sweeping Russia. She managed to receive tutoring in mathematics, at which she excelled. To gain the independence from her family necessary to enable her to pursue further mathematical studies (in German universities, Russian universities being closed to women), she entered at the age of 18 into a fictitious marriage with V. O. Kovalevskii. The ploy proved successful. One year later she had mastered the bureaucratic obstacles posed by the University of Heidelberg and was attending lectures on mathematics and physics there. The following year she moved to Berlin to study with Weierstrass. When her application for admission to the University was denied because of her sex (despite her excellent references), Weierstrass undertook to give her private tutoring. Thus began a warm and enduring friendship between Kovalevskaia and Weierstrass. By the age of 24, Kovalevskaia completed three dissertations for Weierstrass (one being her proof of the Cauchy-Kovalevskaia Theorem). Weierstrass wangled a doctorate for her from Göttingen, the first doctorate in mathematics for a woman since the Renaissance.

Kovalevskaia and her husband, himself newly armed with a doctorate in paleontology, returned to Russia, but failed to find suitable positions, she because of her sex and he because of his arrogance. They were each seduced from their scientific studies by the social and financial whirl of St. Petersburg. In the ensuing years a daughter was born to the Kovalevskii's, Kovalevskaia returned to mathematics, and her husband committed suicide. Her widowhood gave Kovalevskaia both the freedom to pursue her scientific interests and the social respectability necessary to enter the masculine world of the European university. A tenacious campaign by Mittag-Leffler resulted in her appointment first as privat docent at the age of 33 and later as professor at the age of 39 at the University of Stockholm. There she carried out her prize-winning work on the motion of a rigid body about a fixed point. At the age of 41 she succumbed to pneumonia.

Ann Hibner Koblitz presents Kovalevskaia's fascinating biography in a style that is generally lively and graceful. The author is at her best in capturing the social and cultural milieu in which Kovalevskaia moved: Kovalevskaia was on familiar terms with many of the premier mathematicians of the time, her husband and his brother were distinguished scientists, and her path crossed those of Dostoevskii, Chekhov, George Eliot, and Darwin. Koblitz's treatment of Dostoevskii's association with Kovalevskaia's family is particularly noteworthy.

**Kovalevskaia* is a standard English transliteration of the nominative of the feminine form of *Kovalevskii*, her husband's family name (just as *Karenina* is that for *Karenin*). Kovalevskaia, however, published her scientific work in German and French under the transliteration *Kowalevski* of her husband's name. Thus in Latin alphabets her name appears in at least a dozen variants. To complicate matters further, she was affectionately called *Sonia*, the diminutive of *Sofia*.

Throughout her life Kovalevskaia was supremely conscious of her pioneering role as a woman scientist, the foremost of her century. Though endowed with her share of human frailties, she conducted her scientific life with a combination of daring, perserverence, and tactful circumspection. Koblitz carefully documents Kovalevskaia's sympathies for liberalism and for what was then deemed revolutionary. But the hyperbole of the book's title notwithstanding, the author admits (p. 111), "[Kovalevskaia] was not a revolutionary herself."

Koblitz points out that scarcely any of the obstacles Kovalevskaia encountered on account of her sex were erected by mathematicians. Indeed, Königsberger, Hermite, Weierstrass, and especially Mittag-Leffler took great pains to combat the reactionaries opposing her rise to her rightful position.

Most American mathematicians know of Kovalevskaia's life through Bell's [1] superficial portrait, which lacks any bibliographical citation. In contrast, Koblitz's full-scale biography is based upon examination of Russian and Swedish archives and is written with an outward adherence to the norms of historiography. Nevertheless, I found errors of fact and changes of nuance in the interpretation of some quotations from her sources:

(i) The famous episode in which Kovalevskaia persuaded the misogynous chemist Bunsen to admit women to his laboratory is described by Koblitz (p. 89) in a manner totally at variance with that of Bell's amusing tale [1, pp. 424–5], in which Bunsen is portrayed as a harmless crank. The account of Koblitz, citing both Bell and Mittag-Leffler [11, p. 135] is much closer to that of Mittag-Leffler in its lack of theatricality than is Bell's. But Koblitz's statement, "Bunsen... later spread scandalous stories about [Kovalevskaia]" has a flavor different from that of Mittag-Leffler's statement, "[Bunsen] circulait à ce moment des bruits de toutes sortes et non de plus avantageux sur le compte des étudiantes russes qui avaient leur principale résidence à Zuerich." (Although some of Kovalevskaia's Russian friends lived in Zürich, the careful chronology of Koblitz's book gives no indication that she ever did; and she only visited Zürich in 1873.)

(ii) Koblitz asserts that Kovalevskaia was unaware of the efforts of Hermite and Bertrand to rig the Prix Bordin for her: "Kovalevskaia had no way of knowing the behind-the-scenes maneuverings of her French colleagues." (p. 207.) A letter of June 1886 from Kovalevskaia to Mittag-Leffler details how she actually participated in precisely those maneuverings:

> Bertrand always demonstrates towards me an extraordinary benevolence. Imagine what he thought up: next Monday these gentlemen are supposed to gather, to propose a theme for the grand academic prize for the year 1886. Bertrand got the idea of proposing as a theme precisely the problem of rotation of a rigid body. In this way I shall have some chances of getting the prize. You can imagine how much this thought tempts me. Yesterday Hermite, Bertrand, Camille Jordan and Darboux, who are all members of that commission, discussed this project together with me. They made me expose to them once again in detail the results of my work and again heard everything, so that they think, that this work has many chances of being crowned. The only inconvenience is the fact, that I shall have to postpone the publication in that case until 1888. You can imagine how much this project appeals to me. But in that case I shall not be able to make the publication of my work in Christiania in that year... *

(iii) Felix Klein's [4] account of Kovalevskaia's work is branded (along with Bell's) as "unreliable" (p. 279). Koblitz says, "But [Klein] also implies that most of her work was probably done by Weierstrass, and he disparages her later papers," (p. 279). The relevant passage of Klein [4, p. 294] is not so unambiguous: "...ihre Arbeiten in enger Anlehnung und ganz im Stil von Weierstrass geschrieben sind, so dass man nicht sieht, wie weit sie unabhängige, eigene Gedanken enthalten." The first part of this quotation mirrors the view of Mittag-Leffler [10, p. 388]:

*This quotation from [2, p. 223] is apparently an "English" translation of a Russian translation of the French original. I have not been able to see either the original or the Russian translation.

"Comme mathematicien, Sophie Kovalevsky appartient entièrement à l'école de Weierstrass." It is true, however, that Klein did disparage both Kovalevskaia's faulty work on diffraction and her prize-winning work on rigid body rotation, the latter with egregious innuendo: "Ebenso ist man mit der Arbeit über die Rotation nicht durchweg zufrieden." [4, p. 295] What is remarkable about this view, besides the absence of any supporting justification, is that Klein, a coauthor of the monumental treatise on tops [5], was the mathematician of his era best qualified to evaluate Kovalevskaia's work. We shall return to this question below.

In discussing Kovalevskaia's literary works, Koblitz forms her own opinions from her reading of the original Russian versions and expresses them forcefully and authoritatively. But in place of reading Kovalevskaia's mathematical works and related studies, the author relies entirely on others' opinions to evaluate their scientific merit. The progression of favorable quotations, accepted at face value, which form the brief chapter on Kovalevskaia's mathematics soon becomes tiresome. The danger with proof by quotation is apparent from the epigraphs: The first is virtually a tautology, the second might have been true fifty years ago, but is not true now, and the third, couched in the same irresponsible passive voice as used by Klein, is absurd.

Rather than retailing superlatives about Kovalevskaia's genius, it is more satisfying to discuss her mathematical accomplishments. (Koblitz's book, which is addressed to the layman, gives but a superficial account of Kovalevskaia's work.) Kovalevskaia wrote only ten scientific papers, the most important being those devoted to the Cauchy-Kovalevskaia Theorem and to rigid body motion.

The Cauchy-Kovalevskaia Theorem asserts that a partial differential equation involving only analytic functions subjected to analytic initial conditions prescribed on an analytic noncharacteristic initial surface has an analytic solution near that surface that satisfies the initial conditions. On one hand the theorem is general because it applies to equations of all types. On the other hand, it is restrictive because it requires the analyticity of everything in sight, thereby excluding many physical applications, and it is of limited utility because the existence is asserted only on a neighborhood of the initial surface.

Kovalevskaia and Darboux each published proofs of the "Cauchy-Kovalevskaia" Theorem in 1875, whereupon it was discovered that Cauchy had already done so in 1842. Kovalevskaia's proof is the most detailed of the three, those of Cauchy and Darboux appearing in brief notes in the Comptes Rendus. In a letter of recommendation for Kovalevskaia, Hermite commented (p. 241) that Kovalevskaia's paper would be the point of departure for all future research in partial differential equations. Though this theorem has since been invoked from time to time in research work, it has never come near fulfilling Hermite's expectations. Its proof can be found in many books on partial differential equations.

Of more interest is Kovalevskaia's work on the unsymmetrical top, the significance of which has eluded some commentators. To set the stage for a discussion of it, we outline the formulation of the equations of motion of a rigid body about a fixed point taken to be at the origin $\mathbb{0}$ of Euclidean three-space. (Cf. [2], [5], [9].) Let $\mathbb{e}_1(t), \mathbb{e}_2(t), \mathbb{e}_3(t)$ be the principal axes of inertia of the body about $\mathbb{0}$ at time t, which can be taken to be a right-handed orthonormal system fixed in the body. Let I_1, I_2, I_3 be the corresponding principal moments of inertia. Since the body is rigid, I_1, I_2, I_3 are constants. To find the motion of the body we need only find $\mathbb{e}_1, \mathbb{e}_2, \mathbb{e}_3$. The properties of $(\mathbb{e}_1, \mathbb{e}_2, \mathbb{e}_3)$ imply that there is a vector $\omega(t) = \Sigma\omega_j(t)\mathbb{e}_j(t)$, called the *angular velocity* of the body at time t, such that

$$\dot{\mathbb{e}}_j = \omega \times \mathbb{e}_j. \tag{1}$$

Here the superposed dot denotes the time derivative. All summations Σ are taken over $1, 2, 3$. The angular momentum of the body about $\mathbb{0}$ is $\Sigma I_j\omega_j\mathbb{e}_j$. Let $\rho(t) = \Sigma\rho_j\mathbb{e}_j(t)$ be the position of the

mass center of the body at time t. The rigidity of the body implies that ρ_1, ρ_2, ρ_3 are constants. Suppose that the only forces acting on the body are the force of gravity $-mg\mathbf{k}$ and the reaction at the support. Here $\mathbf{k}$ is a fixed unit vector in the vertical direction, g is the acceleration due to gravity, and m is the mass of the body. Suppose that there are no couples applied to the body. (A couple is a pure torque that is not the moment of a force). Then Euler's Law of Motion requiring that the time derivative of the angular momentum about $\mathbf{0}$ equal the resultant torque about $\mathbf{0}$ reduces to

$$\frac{d}{dt}\Sigma I_j\omega_j(t)\mathbf{e}_j(t) = \boldsymbol{\rho}(t)\times[-mg\mathbf{k}]. \tag{2}$$

Let

$$\mathbf{k} = \Sigma k_j(t)\mathbf{e}_j(t). \tag{3}$$

Differentiating this equation with respect to time and using (1) we get

$$\Sigma \dot{k}_j\mathbf{e}_j + \boldsymbol{\omega}\times\mathbf{k} = \mathbf{0}, \tag{4}$$

or equivalently

$$\dot{k}_1 + \omega_2 k_3 - \omega_3 k_2 = 0, \dots, \tag{5}$$

(the ellipses denoting the two equations derived from the exhibited equation by cyclic permutation of the indices). Similarly (2) reduces to

$$I_1\dot{\omega}_1 + (I_3 - I_2)\omega_2\omega_3 = mg(k_2\rho_3 - k_3\rho_2), \dots . \tag{6}$$

Equations (5) and (6) represent a formidable nonlinear sixth-order system of ordinary differential equations for $\omega_1, \omega_2, \omega_3, k_1, k_2, k_3$. (Once these are found, $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ can be determined from additional equations not shown here.)

By dotting (2) with $\boldsymbol{\omega}$ and using (4) we obtain the energy integral:

$$\frac{1}{2}\Sigma I_j\omega_j^2 + mg\boldsymbol{\rho}\cdot\mathbf{k} = \text{constant}. \tag{7}$$

By dotting (2) with $\mathbf{k}$ and using (4) we obtain the conservation of angular momentum about $\mathbf{k}$:

$$\Sigma I_j\omega_j k_j = \text{constant}. \tag{8}$$

Since $\mathbf{k}$ is a unit vector we have a third integral

$$\Sigma k_j^2 = 1. \tag{9}$$

Since t does not appear explicitly in (5) and (6), it can be shown that this system can be reduced to quadratures, i.e., can be solved explicitly in terms of known functions and functions defined by indefinite integrals provided that a fourth integral can be found. There are two important cases in which the fourth integral is readily accessible:

(i) $\boldsymbol{\rho} = \mathbf{0}$ (so that the mass center is at the point of support). Gyroscopes and planets can be described by this case. By multiplying the first equation of (6) by $I_1\omega_1$, the second by $I_2\omega_2$, and the third by $I_3\omega_3$, and then adding the resulting equations we obtain an equation whose integral is $\Sigma I_j^2\omega_j^2 = \text{constant}$.

(ii) $I_1 = I_2, \rho_1 = 0 = \rho_2$. These conditions describe a symmetric top or pendulum. The third equation of (6) immediately yields the integral $\omega_3 = \text{constant}$.

A degenerate case in which the analysis is elementary is

$$I_1 = I_2 = I_3. \tag{iii}$$

The collection of integrals for cases (i) and (ii) can be manipulated algebraically to produce

explicit representations for the triad $\mathbb{e}_1, \mathbb{e}_2, \mathbb{e}_3$ in terms of elliptic functions so dear to the heart of the nineteenth century scientist. To the scientist of the twentieth century the virtue of these integrals is that they yield a simple yet complete qualitative picture of all possible motions, thereby exempting students of rigid body mechanics from the rigors of the theory of elliptic functions.

Being one of the leading experts in the theory of elliptic functions and aware of their role in these simple problems of rigid body motion, Weierstrass proposed (cf. [2, p. 222]) that Kovalevskaia investigate other problems of rigid body motion where her expertise in elliptic functions and more generally in Abelian integrals might be brought to bear. She succeeded in obtaining the fourth integral

$$(10) \qquad \left(\omega_1^2 - \omega_2^2 - mg\rho_1 k_1/I_3\right)^2 + \left(2\omega_1\omega_2 - mg\rho_1 k_2/I_3\right)^2 = \text{constant}.$$

in the case that

$$(\text{iv}) \qquad I_1 = I_2 = 2I_3, \quad \rho_2 = 0 = \rho_3.$$

(The deduction of (10) from (iv) is not obvious, although it is easy to verify that (10) is an integral.) Using (10) Kovalevskaia was able to carry out the full integration of (5) and (6) by an impressively intricate analysis exploiting hyperelliptic integrals. Cf. [2], [9]. For parts of this work she received the Prix Bordin.

Several observations should be made about this work (a) to explain why it is justifiably ignored in many texts on rigid body motion, (b) why it was nevertheless deserving of the Prix Bordin, and (c) how it is influencing some modern work on differential equations.

(a) While conditions (i) and (ii) are frequently encountered in applications, conditions (iv) are very artificial. (The requirement that $\rho_2 = 0$ can be suspended.) This fact and the complicated analysis necessary to solve the problem in terms of special functions have properly led to the relegation of "Kovalevskaia's top" to the rank of a mere curiosity in texts on mechanics.

(b) Conditions (iv) were not a fortuitous result of a search for algebraic integrals. Rather they were the by-product of a systematic analysis of the full system of equations (5), (6). Recall that the solutions in cases (i) and (ii) can be expressed in terms of elliptic functions. When regarded as functions of a complex (time) variable, they have simple poles as their only singularities. Motivated by this observation, Kovalevskaia sought general solutions of (5), (6) that are analytic functions of complex t in a punctured neighborhood of $t = 0$ and that have poles at $t = 0$. These requirements restrict the possible values of the parameters $I_1, I_2, I_3, \rho_1, \rho_2, \rho_3$. Kovalevskaia discovered that cases (i)–(iv) are compatible with her representation. Later Liapunov proved Kovalevskaia's assertion that cases (i)–(iv) are exhaustive. Thus Kovalevskaia's work had the coherence, novelty, and scope necessary to enlist the enthusiastic support of the jury for the Prix Bordin. (Klein's dissatisfaction may have been directed at the gap filled by Liapunov and may further have been influenced by the lack of physical significance of (iv).) Kovalevskaia's work has led to a classification of the elementary cases of (5), (6). Cf. [9, Sects. 7, 8]. (The general problem of rigid body motion is only now beginning to yield to powerful modern methods of analysis. Cf. [3].)

(c) Kovalevskaia's work strongly suggested that one consider the related problem of determining when (5), (6) admits a fourth (independent) integral. According to Leimanis [9, Sect. 7.2], Husson and Burgatti proved that a fourth algebraic integral of (5), (6) exists exactly in the cases (i)–(iv). (If the motion is constrainted, then integrals are known to exist in other cases.) It is remarkable that the purely local and seemingly old-fashioned approach of Kovalevskaia is so intimately related to the question of integrability, which supports modern global qualitative

approaches to the same problem. In recent years it has become evident that completely integrable nonlinear ordinary and partial differential equations, which describe important physical processes, admit very detailed analyses and have solutions with remarkable properties. Simple tests now being developed for determining whether a system is completely integrable can be traced back through the work of Painlevé to works of Fuchs and to Kovalevskaia's local study of the complexified problem of rotation. There still remain, however, deep questions on the relationship of her approach to integrability theory. Cf. **[8]** and the references cited therein. (It is interesting to note that Emmy Noether, Kovalevskaia's successor as the commonly acknowledged greatest female mathematician, made major contributions to the problem of integrability in her 1918 study of invariant variational problems.)

Koblitz's book gives a very readable account of Kovalevskaia's life. It was a life filled with personal drama. The author has uncovered much new material about Kovalevskaia. But the author's admiration for her subject and her lack of expertise in mathematics have led her to color many episodes of Kovalevskaia's mathematical life and assessments of her work in ways more favorable to Kovalevskaia and less favorable to her ostensible opponents than a sober evaluation of the evidence and its context warrants. Kovalevskaia does not require such assistance. She was the author of influential papers and was respected by the mathematical community for mathematical abilities beyond those reflected in her research. She was recognized for her literary efforts. But perhaps her greatest legacy is that today there are many female mathematicians whose *mathematical* accomplishments have surpassed hers.

References

1. E. T. Bell, Men of Mathematics, Simon & Schuster, 1937.

2. Y. V. Golubev, Lectures on Integration of the Equations of Motion of a Rigid Body about a Fixed Point, U.S. Dept. Comm., 1960. (English transl. of Russian edition of 1953.)

3. P. J. Holmes and J. E. Marsden, Horseshoes and Arnold diffusion for Hamiltonian systems on Lie groups, Indiana Univ. Math. J., 32 (1983) 273–309.

4. F. Klein, Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert, Heft I, Springer, 1927.

5. F. Klein and A. Sommerfeld, Über die Theorie des Kreisels, Teubner, 1897–1910.

6. M. Kline, Mathematical Thought from Ancient to Modern Times, Oxford Univ. Press, 1972.

7. E. E. Kramer, Sonya Kovalevsky, Dictionary of Scientific Biography, vol. VII, Scribner's, 1973, pp. 447–480.

8. M. Kruskal, Generalized Painlevé tests for integrability of differential equations, in preparation.

9. E. Leimanis, The General Problem of the Motion of Coupled Rigid Bodies about a Fixed Point, Springer-Verlag, 1965.

10. G. Mittag-Leffler, Sophie Kovalevsky, notice biographique, Acta Math., 16 (1983) 385–392.

11. G. Mittag-Leffler, Weierstrass et Sonja Kowalewsky, Acta Math., 39 (1923) 133–198.

ANSWERS TO "NAME THIS BOOK" ON PAGE 138

1. Roots.
2. Fathers and Sons.
3. Hard Times.
4. The Right Stuff.
5. Westward ho!
6. Far from the Madding Crowd.
7. From Here to Eternity.
8. The End of the Tether.

Carus Mathematical Monograph #21

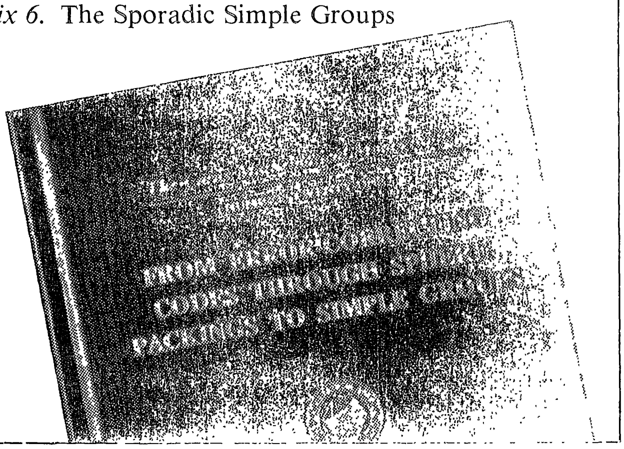
From Error Correcting Codes through Sphere Packings to Simple Groups,

by Thomas M. Thompson
224 pp. Hardbound
List: \$24.00   MAA Member: \$18.50

Two of the most fascinating problems to challenge mathematicians in recent years concern the construction of data transmission codes that can correct errors introduced by static and the search for efficient ways to pack ping-pong balls into a box. Can one design the best error-correcting codes? Can one find the most efficient sphere packing?

Therein lies a fascinating story which is told with great skill and clarity in this important addition to the Carus Mathematical Monograph series. The author has packed (sic) into 175 pages all of the basic mathematical ideas of this saga woven into a gripping historical account of the journey from error-correcting codes to sphere packings to simple groups.

Table of Contents



With D. C. Heath!

COMPUTER SCIENCE

New for 1986

INTRODUCTION TO PASCAL WITH APPLICATIONS IN SCIENCE AND ENGINEERING
Susan Finger, Boston University
Ellen Finger
1986 Paper 720 pages est.
Instructor's Guide/Advanced Applications

Designed for a first course in programming, this strong introduction to Pascal stresses top-down design, modular programming, and problem solving. Ideal for students interested in engineering, mathematics, and science, the text features extensive applications from those areas. Most chapters include a Pascal topic, an application section, and a case study with its algorithmic solution.

INTRODUCTION TO PASCAL AND STRUCTURED DESIGN
Nell Dale, University of Texas at Austin
David Orshalick
Paper 470 pages
Instructor's Guide/Turbo Supplement/ Macintosh Supplement/USCD Supplement

PASCAL PLUS DATA STRUCTURES, ALGORITHMS, AND ADVANCED PROGRAMMING
Nell Dale, University of Texas at Austin
Susan C. Lilly
Paper 490 pages
Instructor's Guide

STRUCTURED COBOL
Gary Haggard,
University of Maine at Orono
Wendell O. Jones,
United States Military Academy
Paper 393 pages
Instructor's Guide

New for 1986

COMPUTERS AND APPLICATIONS: An Introduction To Data Processing
Daniel L. Slotnick, University of Illinois
Evan M. Butterfield, The American University
Ernest S. Colantonio, University of Illinois
Daniel J. Kopetzky, Compion Corporation
Joan K. Slotnick
1986 Cloth/Paper 675 pages
Instructor's Guide/Study Guide/Test Item File/Transparencies/Software Package (Apple® II and IBM®-PC)/Archive Testing Package (Apple® IIe and IBM®-PC)

A comprehensive introduction to computers and data processing, this clearly written text suits readers at all levels of computer literacy. Appropriate for a range of courses, the text is organized for maximum flexibility; its 23 chapters are divided into seven modules which can be taught in almost any order. The text features thorough, accurate treatment of all key computer and data processing concepts.

For details or sample copies, **call us toll free: 800-225-1388.**
In Massachusetts, call collect: 617-863-1948.

Shaping the future since 1885

DC Heath
A Raytheon Company

College Division
125 Spring Street/Lexington, MA 02173

New Mathematical Library #30 . . .

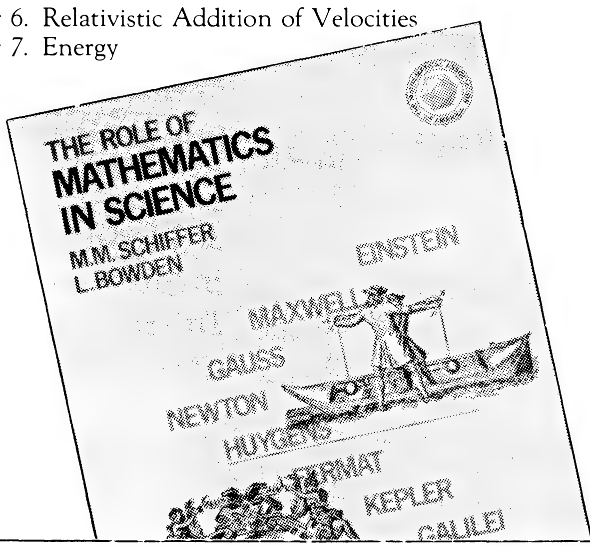
The Role of Mathematics in Science
by Leon Bowden and M.M. Schiffer
220 pp. Paperbound
List: $16.00   MAA Member: $12.50

This book is based on a series of lectures given over a period of years to high school mathematics teachers. The lectures focused on topics in mechanics, questions of population growth, probability and other uses of exponential functions, optics and application of matrices to relativity theory.

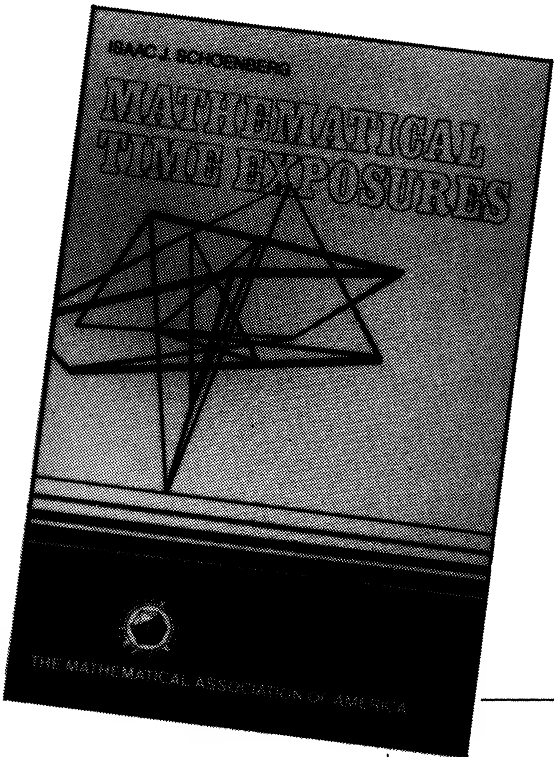
The aim of the book is to illustrate the power and elegance of mathematical reasoning in science with some examples ranging from the work of Archimedes to that of Einstein. The book starts with problems of the lever, the mirror and the growth of populations and ends up with problems of space travel and atomic energy.

This excellent book will appeal to the advanced high school student, to the undergraduate and to their teachers. A sampling of the Table of Contents should inspire you to order your copy now.

Table of Contents



vailable in Hardcover
ınd Paper Editions



Mathematical Time Exposures

by Issac J. Schoenberg
279 pages
Hardcover Edition: List: $34.50
MAA Member: $26.00
Paper Edition List: $21.00
MAA Member: $15.50

Mathematical Time Exposures was inspired by Hugo Steinhaus' admirable book, *Mathematical Snapshots,* published in 1938. The title, *Mathematical Time Exposures* was also suggested by photography, but Schoenberg's pace is much more leisurely than Steinhaus'. Schoenberg spends more time on fewer subjects—the "snapshots" become "time exposures." The subject of at least two of the chapters actually antedate the invention of the daguerreotype.

The author manages to bring together concepts from geometry, number theory, algebra, and analysis, frequently mixing them together in the same chapter. The arts are not neglected. Discussions on the tuning of keyboard instruments, the guitar, and the vibrations of strings are discussed, as well as the suggestion of rectilinear models for outdoor sculpture.

Many of the chapters in this book may be read independently, and at least half of them with only a knowledge of precalculus mathematics.

ORDER FROM:
The Mathematical Association of America
1529 Eioghteenth St, NW
Washington, DC 20036

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 93, Number 3 March 1986

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. (This is necessary for indexing purposes.) Send all main articles, letters to the editor, and editorial correspondence to the editor-elect: H.S. WILF, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Contributions to the other four special departments (Notes, Teaching of Mathematics, Unsolved Problems, and Problems and Solutions) should be sent to the appropriate one of the associate editors, as follows:

Notes: CAROL G. CRAWFORD, Department of Mathematics, U.S. Naval Academy, Annapolis, MD 21402, *or* RICHARD LIBERA, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716.

Teaching of Mathematics: JOAN P. HUTCHINSON *or* STAN WAGON, Department of Mathematics, Smith College, Northampton, MA 01063.

Unsolved Problems: RICHARD GUY, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.

Proposed problems and solutions, both elementary and advanced: PAUL T. BATEMAN, Department of Mathematics, University of Illinois, Urbana, IL 61801.

Reprint Permission: A. B. WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Back Issues: P. and H. BLISS Co., Middletown, CT 06457.

Microfilm Editions: University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

Editorial Assistants: Fanny Hailpern, Mary P. Jackson.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT, ten times a year: January, February, March, April, May, June-July, August-September, October, November and December.

The annual subscription price for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is $24 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $24. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $70 per year.

Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

PRINTED IN THE UNITED STATES OF AMERICA

CUBELIKE PUZZLES—WHAT ARE THEY AND HOW DO YOU SOLVE THEM?

J. A. EIDSWICK
Department of Mathematics and Statistics, University of Nebraska, Lincoln, NE 68588

1. Rubik's cube, Meffert's pyramid, Halpern's dodecahedron, Alexander's great dodecahedron, and the list goes on. Such objects are partitioned into smaller pieces that get mixed up when you turn their faces or layers, yet they don't fall apart. All are groups in disguise; solving them amounts to finding algorithms for factoring arbitrary elements into products of certain generators. Just what constitutes a "cubelike puzzle"? Is there a *universal algorithm* for solving them? These are questions suggested by Douglas Hofstadter's Scientific American article [**12**]. Concerning a puzzle he calls the IncrediBall, Hofstadter writes:

"I found that when I loosened my conceptual grip on the exact qualities of my hard-won operators for the Cube and took them more metaphorically, I could transfer some of my expertise from Cube to I-ball. Not everything transferred, needless to say. What pleased me most was when I discovered that my 'quarkscrew' and 'antiquarkscrew' were directly exportable. Of course, it took a while to determine what such an export would consist of. What is the essence of a move? What aspects of it are provincial and shedable? How can one learn to tell easily? These are difficult questions for which I do not have the answers.

I gradually learned my way around the IncrediBall by realizing that a powerful class of moves consists of turning only two overlapping 'circles' in a commutator pattern ($xyx'y'$). I therefore studied such two-circle commutators on paper until I found ones that filled all my objectives. They included quarkscrews, double swaps and 3-cycles, which form the basis of a complete solution. In doing so I came up with just barely enough notation to cover my needs, but I did not develop a complete notation for the IncrediBall. This, it seems to me, would be most useful: a standard universal notation, psychologically as well as mathematically satisfying, for all cubelike puzzles. It is, however, a very ambitious project, given that you would have to anticipate all conceivable variations on this fertile theme, which is hardly a trivial undertaking."

Concerning algorithms for solving the cube, David Singmaster [**19**, p. 12] writes: "... we need to proceed in two directions. First, by examining the cube and its group, we discover which patterns are possible and, second, we show that we can achieve all possible patterns." The unscrambling problem, then, is directly related to the problem of determining the structure of the underlying group. Later [**19**, pp. 58–9], Singmaster describes this group as a subgroup of index 12 of a direct product of wreath products.

Singmaster also writes: "... it is a remarkable phenomenon that everyone seems to find a different combination of processes and strategy." A survey of the many "how-to" books on the subject will confirm Singmaster's assertion. But common threads will be found running through all of these algorithms. These include the commutators and 3-cycles mentioned by Hofstadter.

This article is an attempt to put some algebraic order into the business of solving cubelike puzzles. Ideally, an algebraic theory would unfold that would, in an elementary way, yield highly efficient algorithms for all such puzzles. The work here is a step in that direction. Our strategy, like Singmaster's, will lead to a determination of the underlying group structures. A description of these structures is given in §6.

Our starting point will be to put the above-mentioned common threads together to form a common strategy. The strategy will then be applied to various puzzles including the general

J. A. Eidswick: I received my Ph.D. in 1964 at Purdue University under the direction of Louis de Branges. Since then I have published modestly in the areas of real analysis and topology. In 1981, I fell under the spell of the cube and wrote the booklet *Rubik's Cube Made Easy*. I also designed *Rubik's Cube Engagement Calendar 1982*. The spell continues. My hobbies include backpacking, gourmet cooking, and jogging enough to justify eating.

$n \times n \times n$ cube. The efficiency of these algorithms will not be a major consideration of this paper, although some attention is given to this topic in §7. It would be very interesting to obtain information about the length of the shortest possible algorithm. Group-theoretically, this amounts to calculating the least upper bound for the lengths of words required to express all elements of the underlying group in terms of a certain set of generators. This is a difficult problem about which little is known. (See [**11**, p. 35], [**19**, pp. 52–3]; also [**20**] for related results.)

In §2, common threads of cube solutions are summarized. In §3, the concept of *wreath product* is developed and examples are given that illustrate the relationship that exists between wreath products and cubelike puzzles. In §4, a general strategy is based on Propositions 1–7. These results serve as keys for solving most cubelike puzzles. Exceptions are the "two faces puzzle" of [**2**, p. 768] (see Example 4 below) and the "skewb" of [**12**, p. 20]. Applications in §5 include the cube puzzle, three different partitions of the tetrahedron, two of the octahedron, two of the dodecahedron, one of the icosahedron, and the general $n \times n \times n$ cube. The latter illustrates the essence of Proposition 1.

The only prerequisite for reading this article is an elementary knowledge of permutation groups. In particular, the reader does *not* have to know how to "do the cube". For a discussion of permutations, see almost any introductory algebra text (e.g., [**3**], [**10**]). For a fairly complete treatment of the subject of permutations and a glimpse at its evolution, see in order, [**5**], [**6**], [**21**], and [**22**]. For the general theory of groups, see, e.g., [**9**], [**15**], or [**16**]. A few "cube theory" references are included at the end for the interested reader.

2. Common threads. All intelligible solutions of the cube puzzle seem to have these common features:

(i) Two distinct subproblems are recognized: the *positioning problem* and the *orientation problem*. Mathematically, these relate to permutation groups and wreath products, respectively.

(ii) Two distinct orbits are recognized: the *corner cube orbit* and the *edge cube orbit*.

(iii) A special *parity-adjusting process* is needed.

(iv) Cubelets are restored *one-by-one*.

(v) Processes for restoring individual cubelets (which involve anywhere from zero to twenty quarter-turns) often involve *conjugations* and *commutators*.

A typical solution begins with operations that affect both orbits, but later restricts to those which affect only one. Likewise, cubelets are positioned and oriented simultaneously at the beginning, then separated later on. There are obviously many ways to do this and therein, no doubt, lies the explanation to Singmaster's observed "phenomenon".

3. Notation, wreath products. Throughout, $\mathscr{X}$ denotes a finite set, G a permutation group acting on $\mathscr{X}$, and Γ a subgroup of a wreath product. In §5, $\mathscr{X}$ will be interpreted as a set of unoriented puzzle pieces, G a group of permutations of such pieces, and Γ a group of permutations of oriented puzzle pieces. Permutations will act on the right, other functions on the left. By an *orbit* of $\mathscr{X}$ is meant a set of the form

$$\mathcal{O} = \mathcal{O}(x) = \{xg \colon g \in G\}.$$

If $\mathcal{O}$ is an orbit, then $G|\mathcal{O}$ denotes the restriction of G to $\mathcal{O}$ and, for $g \in G$, $g|\mathcal{O}$ denotes the restriction of g to $\mathcal{O}$. We will write $\mathrm{Act}(g)$ for the action set $\{x \in \mathscr{X} \colon xg \neq x\}$, $\mathrm{Sym}\,\mathscr{X}$ and $\mathrm{Alt}\,\mathscr{X}$, respectively, for the symmetric and alternating groups on $\mathscr{X}$, $S_n = \mathrm{Sym}\{1,\dots,n\}$, $A_n = \mathrm{Alt}\{1,\dots,n\}$ and $\mathbb{Z}_r$ for the group of integers mod r. If S is a finite set, $|S|$ denotes its cardinality and $S^{\mathscr{X}}$ the set of functions from $\mathscr{X}$ to S. If g and h are elements of a group, then $[g,h]$ denotes the *commutator* $ghg^{-1}h^{-1}$.

Wreath products are usually studied along with *semidirect products* and/or *group extensions* as, e.g., in [**13**], [**15**], and [**18**]. The definition below (cf. [**14**, p. 32]) assumes no previous knowledge of these companion ideas. We mention, though, the important theorem of Kaloujnine and Krasner

(see [**18**, p. 100] or [**13**, p. 49]): *Every extension of a group A by a group B can be embedded in a wreath product.*

DEFINITION. Let G and H be permutation groups that act on $\mathscr{X} = \{1,\dots,n\}$ and $\mathscr{Y} = \{1,\dots,r\}$, respectively. Then the *wreath product of H by G*, written $H \wr G$, is the subgroup of $\mathrm{Sym}(\mathscr{X} \times \mathscr{Y})$ generated by permutations of the following two types:

$$\pi(g) : (i, j) \to (ig, j)$$

for $g \in G$ and

$$\sigma(h_1,\dots,h_n) : (i, j) \to (i, jh_i)$$

for $h_1,\dots,h_n \in H$.

We may visualize the situation as follows: Suppose n decks of cards occupy positions $1,\dots,n$ and suppose that each deck contains r cards which occupy levels $1,\dots,r$. Then $\pi(g)$ permutes the decks according to g, maintaining card levels, and $\sigma(h_1,\dots,h_n)$ shuffles the decks in positions $1,\dots,n$ according to $h_1,\dots,h_n$, respectively, maintaining deck positions.

In the applications to be considered, H will be cyclic. Accordingly, we will take $H = \mathbb{Z}_r$ for some positive integer r and use additive notation for this group. Notice that for such groups, shufflings amount to what card players call "cuts".

For $g, p \in G$ and $h_1,\dots,h_n, k_1,\dots,k_n \in H$, we have

$$\pi(g)\pi(p) = \pi(gp),$$
$$\sigma(h_1,\dots,h_n)\sigma(k_1,\dots,k_n) = \sigma(h_1 + k_1,\dots,h_n + k_n),$$

and

$$\sigma(h_1,\dots,h_n)\pi(g) = \pi(g)\sigma\left(h_{1g^{-1}},\dots,h_{ng^{-1}}\right).$$

It follows that any element α of $H \wr G$ has a unique representation of the form $\pi(\alpha')\sigma(\alpha'')$, where $\alpha' \in G$ and $\alpha'' \in H^{\mathscr{X}}$. In particular, we have

$$|H \wr G| = |G \times H^{\mathscr{X}}| = |G||H|^{|\mathscr{X}|}.$$

Henceforth, we will identify elements α of $H \wr G$ and corresponding pairs (α', α'') of $G \times H^{\mathscr{X}}$. This will simplify notation. For example, the identity 1 in $H \wr G$ will be written $(1,0)$ or $(1,(0,\dots,0))$ instead of $\pi(1)\sigma(0,\dots,0)$. The following formulas may be routinely established:

$$(\alpha\beta)' = \alpha'\beta' \tag{1}$$

$$(\alpha^{-1})' = \alpha'^{-1} \tag{2}$$

$$\left(\alpha\beta\alpha^{-1}\right)' = \alpha'\beta'\alpha'^{-1} \tag{3}$$

$$[\alpha,\beta]' = [\alpha',\beta'] \tag{4}$$

$$(\alpha\beta)''(i) = \alpha''\left(i\beta'^{-1}\right) + \beta''(i) \tag{5}$$

$$(\alpha^{-1})''(i) = -\alpha''(i\alpha') \tag{6}$$

$$\left(\alpha\beta\alpha^{-1}\right)''(i) = \alpha''\left(i\alpha'\beta'^{-1}\right) + \beta''(i\alpha') - \alpha''(i\alpha') \tag{7}$$

$$[\alpha,\beta]''(i) = \alpha''\left(i\beta'\alpha'\beta'^{-1}\right) + \beta''(i\beta'\alpha') - \alpha''(i\beta'\alpha') - \beta''(i\beta'). \tag{8}$$

The following three examples show how wreath products relate to the puzzles under consideration. They also illustrate the abovementioned theorem of Kaloujnine and Krasner. The first example is a variation of "the three coins game" of [**4**, p. 24].

EXAMPLE 1. Three coins are placed in a row in positions 1, 2, 3 and the following operations on

them are permitted.

α: interchange the coins in positions 1 and 2.
β: turn over the coin in position 3, then interchange coins in positions 2 and 3.

We assume that the coins have distinguishable orientations "heads" and "tails" and (unlike in [**4**]) that the coins themselves are distinguishable (say, penny, nickel, and dime). If we associate coin positions with deck positions and "heads" and "tails", respectively, with levels 1 and 2, it is clear that α and β may be regarded as permuting-shuffling operations on three 2-card decks; i.e., as elements of $\mathbb{Z}_2 \wr S_3$. Explicitly we have:

$$\alpha = ((12),(0,0,0))$$

and

$$\beta = ((23),(0,1,0)).$$

One may verify by direct calculation that α and β generate all 48 elements of $\mathbb{Z}_2 \wr S_3$ (however, compare this with Examples 2 and 3 below).

EXAMPLE 2. Six equilateral triangles are placed in a row in positions $1,\ldots,6$ and the following operations on them are allowed.

α: cycle the triangles in positions 3, 4, 5, 6 in that order (i.e., $3 \to 4 \to 5 \to 6 \to 3$).
β: rotate the triangle in position 1 clockwise 120°, then cycle the triangles in positions $1,2,3,4$ in that order (i.e., $1 \to 2 \to 3 \to 4 \to 1$).

(It is to be understood here that when triangles are cycled, they are to maintain their orientations. Thus, vertices pointing upward before cycling will remain pointed upward after cycling.) If we identify the idea of having i clockwise rotations and that of occupying level i (mod 3), then α and β may be interpreted as permuting-shuffling operations on six 3-card decks. Hence, the group generated by α and β may be regarded as a subgroup Γ of $\mathbb{Z}_3 \wr S_6$. Here,

$$\alpha = ((3456),(0,0,0,0,0,0))$$

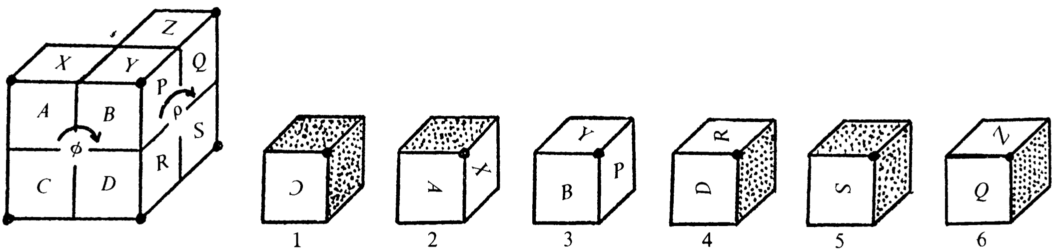
and

$$\beta = ((1234),(0,1,0,0,0,0)),$$

and it may be verified that Γ has index 6 in $\mathbb{Z}_3 \wr S_6$. Indeed, an argument similar to [**19**, pp. 55–7] will show that Γ is isomorphic to $\mathbb{Z}_3 \wr S_5$. This example is related to the "Tricky Six Puzzle" of [**2**, p. 759] (see also [**19**, p. 57], [**23**], and Example 3 below).

EXAMPLE 3 (*the* 2-*faces group*). Six blocks are arranged in an L-shaped stack as shown in Fig. 1(a), and the following operations are allowed (see Fig. 1(a)).

ϕ: rotate the front layer 90° clockwise.
ρ: rotate the right-hand layer 90° clockwise.



(a) (b)

FIG. 1

This example is closely related to Example 2. To see this, note that ϕ and ρ take darkened corners to darkened corners (see Fig. 1(a)), and that blocks can rotate about these corners (e.g., $\phi\rho$: $B \to Y \to P \to B$). Now imagine the six blocks as spread out before us in some arrangement, say, as in Fig. 1(b). We may then view ϕ and ρ as permuting-rotating operations as in Example 2. Clearly, the representations of ϕ and ρ will depend on the arrangement, but the point is that we may regard the group Γ generated by ϕ and ρ as a subgroup of $\mathbb{Z}_3 \wr S_6$. In terms of the arrangement shown,

$$\phi = ((1234), (0,0,0,0,0,0))$$

and

$$\rho = ((3654), (0,0,1,1,0,1)).$$

It may be verified (see [**19**, pp. 55–7]) that Γ has index 18 in $\mathbb{Z}_3 \wr S_6$.

4. A common strategy. In the puzzles we are considering there are various pieces which can occupy certain positions with certain orientations. The basic moves of the puzzle may be viewed as acting on both the positions and the oriented positions. When we view these moves as acting on positions, we will obtain orbits $\mathcal{O}_1, \ldots, \mathcal{O}_N$. When we view the moves as acting on the oriented positions restricted to an orbit $\mathcal{O}_i$, the situation will be analogous to Example 3. Thus, these puzzles may be regarded as subdirect products of wreath products of the form $\mathbb{Z}_r \wr S_n$.

Our strategy will be to proceed through the orbits $\mathcal{O}_1, \ldots, \mathcal{O}_N$ restoring those pieces in a given orbit without disturbing the other pieces. To accomplish this, we will first need to make a parity adjustment. Let α be the product of moves that scrambled up our puzzle. We will produce a product π of basic moves such that $\alpha\pi$ permutes the pieces *evenly* on every orbit. We will then produce products $\alpha_1, \ldots, \alpha_N$ of basic moves such that α_i acts on $\mathcal{O}_j$ as the identity for $j \neq i$, and α_i acts on $\mathcal{O}_i$ as the inverse of $\alpha\pi$. (In fact, we need only require that α_i be the identity on $\mathcal{O}_j$ for $j < i$ and that α_i act on $\mathcal{O}_i$ as the inverse of $\alpha\pi\alpha_1 \cdots \alpha_{i-1}$. This improvement to the strategy will be discussed in §7.) Thus, the product $\alpha\pi\alpha_1 \cdots \alpha_N$ will unscramble the puzzle.

The following result shows that a parity adjustment of the type described above is always possible. We define $\nu(g, \mathcal{O})$ to be 0 or 1 according as $g|\mathcal{O}$ is even or odd.

PROPOSITION 1. *For any permutation group G there exist orbits $\mathcal{O}_1, \ldots, \mathcal{O}_m$ and elements $h_1, \ldots, h_m \in G$ such that*

(i) *$\nu(h_i, \mathcal{O}_j) = \delta_{ij}$ $(i, j = 1, \ldots, m)$ and*

(ii) *for any orbit $\mathcal{O}$ there exist $c_1, \ldots, c_m \in \{0,1\}$ such that*

$$\nu(h, \mathcal{O}) = \sum_{i=1}^{m} c_i \nu(h, \mathcal{O}_i)$$

mod 2 *for every $h \in G$.*

Moreover, if $\mathcal{O}_1, \ldots, \mathcal{O}_m$ and $h_1, \ldots, h_m$ are as above, then

(iii) *for any $g \in G$, $\nu(gp, \mathcal{O}) = 0$ for every orbit $\mathcal{O}$ where p is the product of the h_i's, taken in order, for which $\nu(g, \mathcal{O}_i) = 1$.*

Proof. The following argument, contributed by T. Shores, is "programmable". Let $G = \{g_1, \ldots, g_k\}$, let $\{\mathcal{X}_1, \ldots, \mathcal{X}_l\}$ be the set of orbits of G, and consider the $k \times l$ matrix $M = [\nu(g_i, \mathcal{X}_j)]$. Since the mapping $g \to [\nu(g, \mathcal{X}_j)]$ is a homomorphism from G into $Z_2^{(l)}$, the row-reduced echelon matrix E of M has the form: $E = [\nu(g_{r_i}, \mathcal{X}_j)]$. If E has m nonzero rows, let $t_1, \ldots, t_m$ be the columns corresponding to the leading 1's in these rows. Then (i) and (ii) follow with $h_i = g_{r_i}$ and $\mathcal{O}_i = \mathcal{X}_{t_i}$. Also for g and p as in (iii), we have $\nu(gp, \mathcal{O}_i) = 0$ for $i = 1, \ldots, m$ and hence $\nu(gp, \mathcal{O}) = 0$ for every orbit $\mathcal{O}$ by (ii). ∎

Since the strategy outlined above deals with each orbit separately, we will concentrate for the remainder of this section on the action of a wreath product on a single orbit.

Recall that for $\alpha \in H \wr G$, α' and α'' denote the "permuting" and "shuffling" parts of α, respectively, i.e., $\alpha = (\alpha', \alpha'')$ where $\alpha' \in G$ and $\alpha'' \in H^{\mathscr{X}}$. We will also write $\text{Act}'(\alpha)$ for the projection of $\text{Act}(\alpha)$ on $\mathscr{X}$. Thus,

$$\text{Act}'(\alpha) = \{ i \in \mathscr{X} \colon i\alpha' \neq i \quad \text{or} \quad \alpha''(i) \neq 0 \}.$$

An element $\alpha \in H \wr G$ will be called a k-cycle if α' is a k-cycle, $\text{Act}'(\alpha) = \text{Act}(\alpha')$, and α has order k. Cubelike puzzles are generated by products of such cycles. *Examples*: (i) ϕ and ρ of Example 3 are 4-cycles. (ii) β of Example 1 [Example 2] is not a cycle because β' is a 2-cycle [4-cycle] and the order of β is 4 [12].

By formula (5),

$$(\alpha^k)''(i) = \sum_{j=0}^{k-1} \alpha''(i\alpha'^{-j}) \quad \text{for} \quad i = 1, \dots, n.$$

Hence, it follows that

$$\sum_{i=1}^{n} \alpha''(i) = 0 \tag{9}$$

holds for any cycle α. Since property (9) is clearly preserved under multiplication in a wreath product, Proposition 2 below follows. The result, almost trivial in this context, translates into puzzle "conservation laws" like "the number of flipped cubies is always even" and "total clockwise twist equals total counterclockwise twist mod 3". Elaborate arguments have been given to explain these phenomena. See, e.g., [**11**, p. 28], [**19**, p. 17], and [**2**, p. 762].

PROPOSITION 2. *If Γ is a subgroup of $H \wr G$ which is generated by cycles, then* (9) *holds for all* $\alpha \in \Gamma$.

By a *complete set of* 3-*cycles for* $a, b, c \in \mathscr{X}$ will be meant a set $\{\tau_0, \dots, \tau_{r-1}\}$ of 3-cycles such that for each $j = 0, 1, \dots, r-1$,

$$\tau_j' = (abc) \tag{i}$$

and

$$\tau_j''(a) = \tau_0''(a) + j, \qquad \tau_j''(b) = \tau_0''(b), \qquad \tau_j''(c) = \tau_0''(c) - j. \tag{ii}$$

The following result shows how to produce 3-cycles.

PROPOSITION 3. *If α, β and σ satisfy*

(i) $\text{Act}'(\alpha) \cap \text{Act}'(\beta) = \{b\}$, *where $a\alpha' = b = c\beta'$, and*

(ii) $\text{Act}'(\sigma) \cap \{a, b, c\} = \{c\}$, *where $c\sigma' = c$ and $\sigma''(c) = 1$,*

then a complete set of 3-*cycles for a, b, c is given by*

$$\tau_j = \sigma^j [\alpha, \beta] \sigma^{-j}$$

for $j = 0, 1, \dots, r-1$.

Proof. Let $\tau_0 = [\alpha, \beta]$. By formulas (4) and (8) and condition (i),

$$\tau_0' = [\alpha', \beta'] = (abc), \quad \tau_0''(a) = \beta''(b) - \alpha''(b),$$

$$\tau_0''(b) = \alpha''(b), \quad \tau_0''(c) = -\beta''(b), \quad \text{and} \quad \tau_0''(i) = 0 \quad \text{for } i \neq a, b, c.$$

Hence, τ_0 is a 3-cycle. Let $\tau_1 = \sigma\tau_0\sigma^{-1}$. Then by formulas (3) and (7) and conditions (i) and (ii),

$$\tau_1' = \sigma'(abc)\sigma'^{-1} = (abc), \quad \tau_1''(a) = \tau_0''(a) + 1,$$

$$\tau_1''(b) = \tau_0''(b), \quad \tau_1''(c) = \tau_0''(c) - 1, \quad \text{and} \quad \tau_1''(i) = 0 \quad \text{for } i \neq a, b, c.$$

The result follows by induction. ∎

Elements of the form $\mu(a,b) = (1,s)$, where $s(a) = 1$, $s(b) = -1$, and $s(i) = 0$ otherwise, play a special role in cubelike puzzles. For the case of the corner orbit of the cube puzzle, these elements are popularly called *mesons* because of an interesting link with particle physics (see [**8**]). Since $\tau_1\tau_0^{-1} = \mu(c,b)$, Proposition 3 also shows us how to produce these generalized mesons.

COROLLARY. *Under the assumptions of Proposition* 3, *we have*

$$\mu(c,b) = [\sigma,[\alpha,\beta]].$$

Typically, Proposition 3 will be applied with β having one of these two forms: $\beta = \gamma\delta\gamma^{-1}$ or $\beta = [\gamma,\delta]$. Propositions 4 and 5 below show how to deal with these special situations.

PROPOSITION 4. *If* α, γ, *and* δ *satisfy*

$$\mathrm{Act}'(\alpha)\gamma' \cap \mathrm{Act}'(\delta) = \{b\gamma'\},$$

then $Act'(\alpha) \cap Act'(\gamma\delta\gamma^{-1}) = \{b\}$.

Proof. By (3) and (7), $i \in \mathrm{Act}'(\beta\gamma\beta^{-1})$ if and only if $i\beta' \in \mathrm{Act}'(\gamma)$ and the result follows. ■

PROPOSITION 5. *If* α, γ, *and* δ *satisfy*
(i) $\mathrm{Act}'(\alpha) \cap \mathrm{Act}'(\gamma) \cap \mathrm{Act}'(\delta) = \{b\}$
(ii) $(\mathrm{Act}'(\alpha) \cap \mathrm{Act}'(\gamma))\gamma' \cap \mathrm{Act}'(\delta) = \varnothing$, *and*
(iii) $(\mathrm{Act}'(\alpha) \cap \mathrm{Act}'(\delta))\delta' \cap \mathrm{Act}'(\gamma) = \varnothing$,
then $\mathrm{Act}'(\alpha) \cap \mathrm{Act}'([\gamma,\delta]) = \{b\}$.

Proof. By hypothesis, $b\gamma'\delta' = b\gamma' \neq b\delta' = b\delta'\gamma'$; hence, $b \in \mathrm{Act}([\gamma',\delta'])$. Conversely, if $i \notin \mathrm{Act}'(\gamma)$, then $i \in \mathrm{Act}'([\gamma,\delta])$ if and only if $i\delta' \in \mathrm{Act}'(\gamma)$. Therefore, if $i \in \mathrm{Act}'(\alpha) \cap \mathrm{Act}'([\gamma,\delta])$, then by (iii), we must have $i \in \mathrm{Act}'(\gamma)$. Similarly, by (ii), $i \in \mathrm{Act}'(\delta)$ and so $i = b$ by (i). ■

For all of the puzzles to be considered the group G will be triply transitive and the following condition easy to verify.

PROPOSITION 6. *Let* Γ *be a subgroup of* $H \wr G$. *If* Γ *contains a complete set of* 3*-cycles for three elements* a, b, c *and if for each* $x \in \mathscr{X}$, $x \neq a, b, c$, *there exists* $\alpha \in \Gamma$ *such that*

$$\mathrm{Act}(\alpha') \cap \{a,b,c\} = \{x\alpha'\},$$

then Γ *contains a complete set of* 3*-cycles for any three elements of* $\mathscr{X}$.

Proof. From the proof of Proposition 3, we see that conjugation of a complete set of 3-cycles yields a complete set of 3-cycles. Hence, to prove the result we need only express 3-cycles of $\mathscr{X}$ as conjugates of (abc). For $x \in \mathscr{X}$, $x \neq a, b, c$, assume

$$\mathrm{Act}(p) \cap \{a,b,c\} = \{a\} = \{xp\}.$$

Then

$$(xbc) = p(abc)p^{-1},$$
$$(axc) = (abc)^{-1}(xbc)(abc),$$

and

$$(abx) = (abc)(xbc)(abc)^{-1}.$$

From these we get all 3-cycles as follows:

$$(xyc) = (aby)^{-1}(xbc)(aby),$$
$$(xby) = (axc)^{-1}(aby)(axc),$$
$$(axy) = (ybc)^{-1}(axc)(ybc),$$

and

$$(xyz) = (zbc)^{-1}(xyc)(zbc). \blacksquare$$

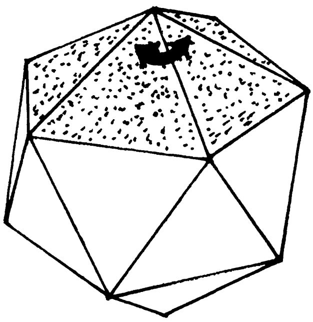
PROPOSITION 7. *If Γ is a subgroup of $H \wr G$ that contains a complete set of 3-cycles for any three elements of $\mathscr{X}$, then Γ contains all elements α such that α' is even and α'' satisfies* (9). *For any such α, there exist 3-cycles $\tau_1, \dots, \tau_N \in \Gamma$ such that $\alpha\tau_1 \cdots \tau_N = 1$.*

Proof. The proof is by induction on the number of elements of $\mathrm{Act}'(\alpha)$. If $\alpha' \neq 1$, there exist distinct elements a, b, and c of $\mathrm{Act}(\alpha')$ with $b = a\alpha'$. Let τ be a 3-cycle such that $\tau' = (acb)$ and $\tau''(a) = -\alpha''(b)$. Then $|\mathrm{Act}'(\alpha\tau)| < |\mathrm{Act}'(\alpha)|$. If $\alpha' = 1$, there exist distinct elements a and b with $\alpha''(a) \neq 0$ and $\alpha''(b) \neq 0$. Then $|\mathrm{Act}'(\alpha\mu(a,b)^{\alpha''(b)})| < |\mathrm{Act}'(\alpha)|$. $\blacksquare$

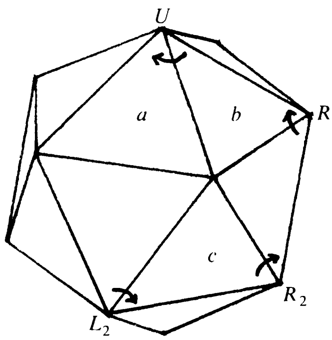
5. Applications. This section contains various applications. Discussion is limited to puzzles based on partitioned regular polyhedra. It is assumed that puzzle pieces are marked so that home locations and orientations are uniquely determined. To keep the notation manageable, in most cases only those generators and puzzle pieces that play direct roles in the illustrated moves will be given names. The reader may find it helpful at times to express generators as products of k-cycles using a more complete notational system. A complete notational system will be developed for the $n \times n \times n$ cube puzzle.

Except for slight variations, the devices considered here actually exist. The $n \times n \times n$ cube exists for $n = 2, 3, 4$, and 5, and there is even a transparent $5 \times 5 \times 5$ cube which is held together by magnets.

The icosahedron puzzle. We start with this example because it is simplest and displays many of the basic ideas. The generators are the twelve 72° vertex turns as indicated in Fig. 2(a). Products of these generators cause permutations of the twenty triangular faces. They also cause rotations as can be seen by studying the action of the move L_2R_2 (see Fig. 2(b)) on the vertices of the face marked c. By the discussion in §3, we see that the combined permuting-rotating action can be regarded as taking place in the wreath product $\mathbb{Z}_3 \wr S_{20}$.



(a)



(b)

FIG. 2

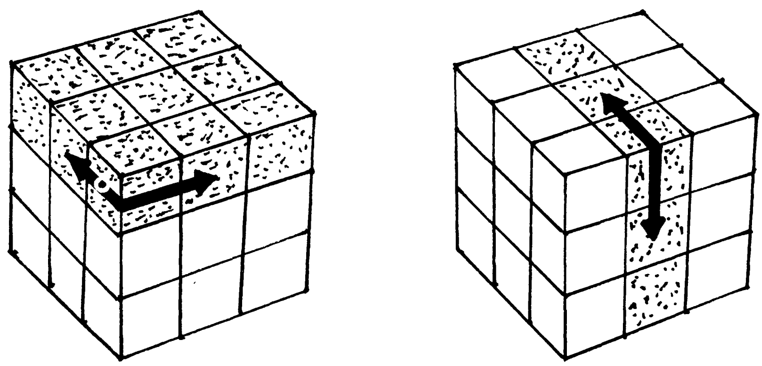
Now let's see how to cycle the pieces marked a, b, c in Fig. 2(b). To apply Proposition 3 we need "disjoint" moves α and β such that α': $a \to b$ and β': $c \to b$. That U^{-1}, $R_1^{-1}R_2R_1$ is such a pair can be seen from Fig. 2(b) and can also be verified by Proposition 4. Since L_2R_2 causes a 120° clockwise rotation of piece c and stabilizes a and b, it follows that a complete set of 3-cycles for a, b, c is given by:

$$(abc)_j = (L_2R_2)^j\left[U^{-1}, R_1^{-1}R_2R_1\right](L_2R_2)^{-j} \quad (j = 0, \pm 1).$$

We can regard this set as a vehicle for moving piece b into location c ending up with any

orientation we please. Note that the move corresponding to $j = 0$ affects the orientation in the same way as the move $R_1^{-1}R_2^{-1}$ and the move corresponding to $j = -1$ $[j = 1]$ adds a 120° clockwise [counterclockwise] rotation.

Since the generators for this puzzle are 5-cycles, all patterns correspond to even permutations and Proposition 1 plays no role. Clearly, Proposition 6 (and its proof) apply and we may proceed as in Proposition 7 to unscramble any possible configuration.



(a) (b)

FIG. 3

The cube puzzle. This puzzle can be viewed either as being generated by the six 90° face turns indicated in Fig. 3(a) or by the nine 90° layer turns which include the three "slice turns" indicated in Fig. 3(b). Since a slice turn is clearly equivalent to a pair of parallel face turns followed by a 90° rotation of the entire cube, no new patterns arise by allowing these moves. Group-theoretically, if M_f and M_l are the groups generated by faces and layers, respectively, then M_l is an extension of M_f by the group of rigid motions of the cube.

We will work in the larger group, allowing slice turns, because it will lead to neater formulas. The price for this convenience will be the inconvenience of having to deal with a trivial orbit. The orbits are: $\mathscr{V}$ consisting of the eight corner positions, $\mathscr{E}$ consisting of the twelve edge positions, and $\mathscr{C}$, the trivial orbit, consisting of the six face centers. Corresponding to elements of $\mathscr{V}$, $\mathscr{E}$, and $\mathscr{C}$, respectively, there are 3, 2, and 4 possible orientations. We note, however, that in the most popular version of this puzzle, the six faces are solidly colored rendering center piece rotations indistinguishable.

Let h_1 be any face turn, h_2 any slice turn, and let $\mathscr{O}_1 = \mathscr{V}$, $\mathscr{O}_2 = \mathscr{C}$. Then, for $m = 2$, (i) of Proposition 1 is clearly satisfied. Moreover, if h is any generator, (ii) of Proposition 1 is satisfied by $\mathscr{O} = \mathscr{E}$ and $c_1 = c_2 = 1$. It follows that (ii) holds for all products of generators, i.e., for all moves h. The conclusion (iii) of Proposition 1 says the obvious in this case: Any element can be made even on all orbits by making at most two quarter-turns. Note that if face centers are correctly positioned, at most one quarter-turn is needed, and, if, in addition, face centers are correctly oriented, then no adjustment is needed.

As with the icosahedron puzzle, Propositions 3–4 give us the following set of 3-cycles, Proposition 6 applies to both orbits $\mathscr{V}$ and $\mathscr{E}$ and we can proceed as in Proposition 7. The notation refers to Fig. 4.

$$(abc)_j = (LD)^j[U, R^{-1}DR](LD)^{-j} \quad (j = 0, \pm 1),$$

$$(123)_j = (\rho^{-1}F^2)^j[\phi^{-1}, R^{-1}FR](\rho^{-1}F^2)^{-j} \quad (j = 0, 1).$$

Except for a possible parity adjustment at the beginning, the solution is a product of moves of the form $\gamma'[\alpha, \beta]\gamma^{-'}$ which stabilize the oriented centers. It follows that if the center squares are

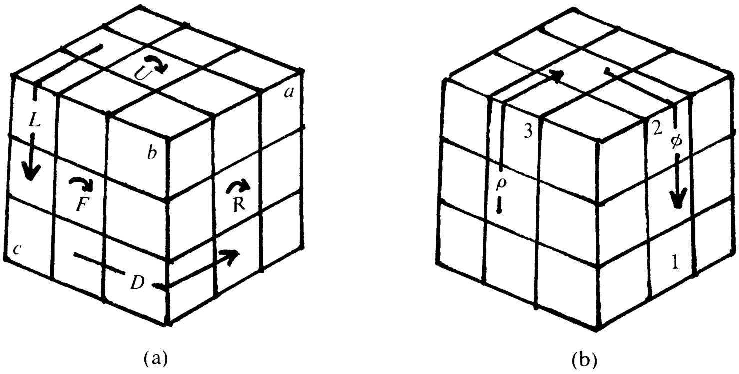


FIG. 4

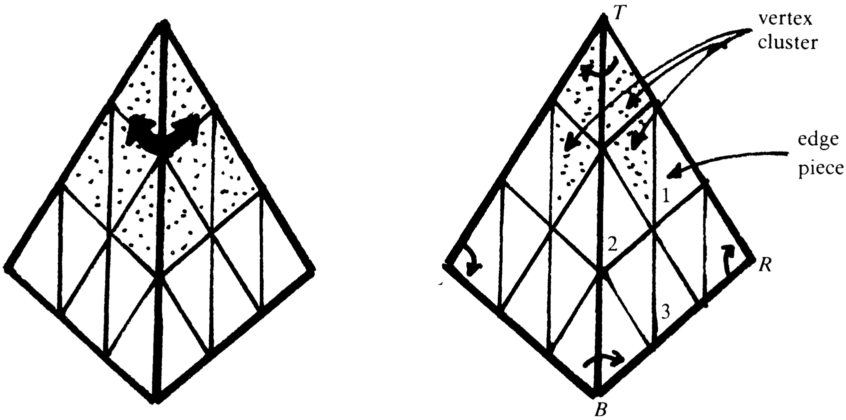


FIG. 5

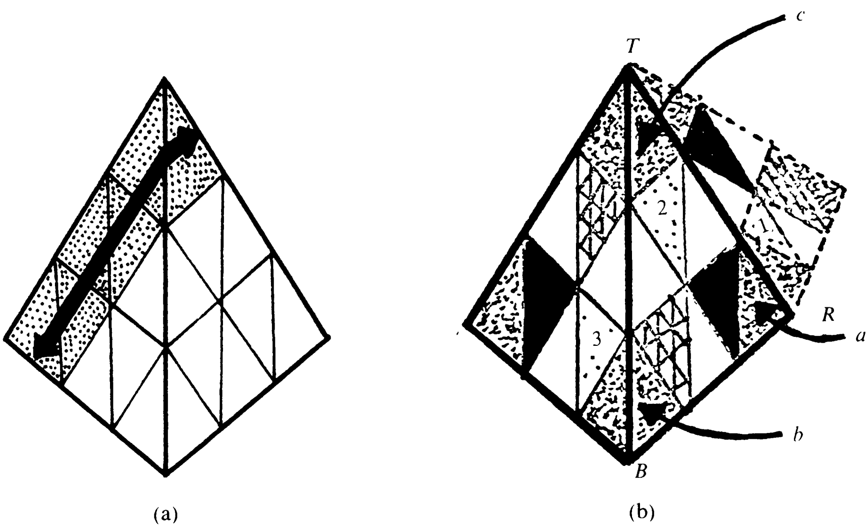


FIG. 6

correctly positioned and oriented at the beginning, they will end up that way.

Tetrahedron puzzles. Two versions will be discussed. When combined, a third version (called the Master Pyraminx in [**12**]) results which can be solved by combining the two solutions.

The first puzzle is generated by the four 120° vertex turns as indicated in Fig. 5. Note that each vertex consists of a cluster of four pieces which rotates as a unit independent of the other vertex clusters. (For the puzzle called the Popular Pyraminx in [**12**] the little tetrahedron tips can also be rotated. This complication is trivial and is disregarded here.) The vertex clusters play the same pivotal role as the center squares for the cube puzzle without slices. As with the cube puzzle, it will be seen that they can be fixed up at the outset and never again worried about. Thus, there is only one orbit which consists of the six edge pieces (see Fig. 5). As before, the following complete set of 3-cycles can be obtained via Propositions 3–4, Proposition 6 applies, and we can proceed as in Proposition 7. The notation refers to Fig. 5:

$$(123)_j = (RTBT^{-1})^j[T, B^{-1}](RTBT^{-1})^{-j} \quad (j = 0,1).$$

The second version is generated by the 180° edge turns as indicated in Fig. 6(a). There are four orbits: $\mathscr{V}$ consisting of the four corner pieces and three face orbits $\mathscr{F}_1, \mathscr{F}_2, \mathscr{F}_3$ each consisting of four face pieces. The orbits are indicated in Fig. 6(b) by different shadings.

Note that the six edge pieces remain fixed relative to one another and play the same pivotal role as the center squares in the cube puzzle and the vertex clusters in the preceding example. Clearly,

$$\nu(g, \mathscr{F}_1) + \nu(g, \mathscr{F}_2) + \nu(g, \mathscr{F}_3) = 0$$

mod 2 for any generator g and hence for all g, and Proposition 1 is satisfied with $m = 3$. Note, though, that parities are taken care of automatically by correctly orienting the edge pieces. Note also that orientations of the remaining pieces are uniquely determined.

By Propositions 3–4, we obtain:

$$(123) = [\overline{RT}, \overline{LT}\,\overline{BT}\,\overline{LT}],$$

where the notation refers to Fig. 6 and $\overline{V_1V_2}$ denotes the turn with ends V_1, V_2. Similar 3-cycles can be found for the other two face orbits.

It is easy to see that the hypothesis of Proposition 5 is satisfied with $\alpha = \overline{BR}$, $\gamma = \overline{BT}$, and $\delta = \overline{BL}$, and, hence,

$$(abc) = [\overline{BR}, [\overline{BT}, \overline{BL}]].$$

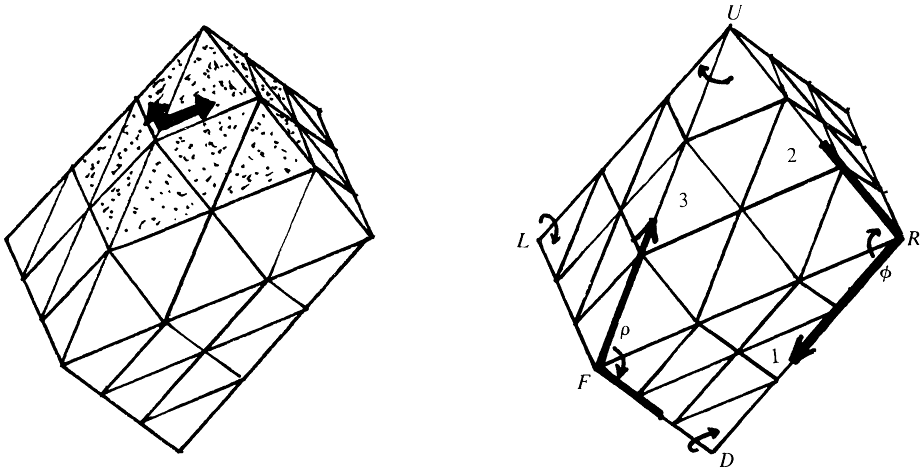
Proposition 6 clearly applies to all orbits and the solution can be completed via Proposition 7.

Octahedron puzzles. Two versions will be discussed. The first is generated by the six 90° vertex turns as indicated in Fig. 7(a). As pointed out in [**12**] there is a vertex/center duality between this puzzle and the cube puzzle. The octahedron is easier because the eight corner cubelets of the cube puzzle correspond to mere points on the octahedron puzzle. Thus, a simple strategy is to first orient the vertex clusters, then use edge processes developed for the cube puzzle (compare Figs. 4 and 7(b)):

$$(123)_j = (\rho^{-1}F^2)^j[\phi^{-1}, R^{-1}FR](\rho^{-1}F^2)^{-j} \quad (j = 0,1).$$

The second puzzle is the face-turning puzzle described in [**12**]. Like the cube puzzle, we will allow slice turns (see Fig. 9). The solution given here provides for the addition of orientable pieces at the face centers, the idea being that such a center piece would rotate only when its corresponding face rotated. See Fig. 8. The puzzle without orientable center pieces is somewhat easier.

There are four orbits: $\mathscr{V}$ consisting of the six corners, $\mathscr{E}$ consisting of the twelve edges, and two face orbits $\mathscr{F}_1$ and $\mathscr{F}_2$ each consisting of twelve pieces. The orbits are indicated in Fig. 9 and



(a) (b)

FIG. 7

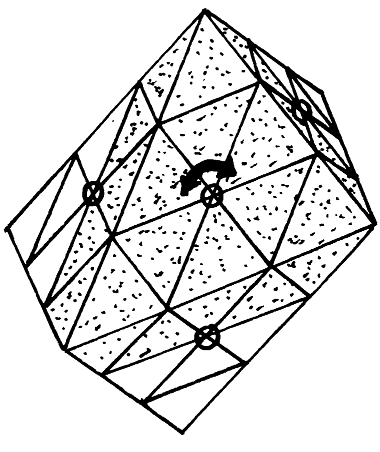


FIG. 8

again in Fig. 10, which is a flattened version of the puzzle. In the flattened version, one can readily view the entire action of the generators. Of course, these can all be expressed as products of 3-cycles (five for face turns, four for slice turns) and the reader, in checking the formulas below, may well find it advantageous to do so.

Observe that edge and face pieces have only one possible orientation and that corners have two. By Propositions 3–4,

$$(e_1e_2e_3) = [F_2, \rho_1^{-1}\phi_2\rho_1],$$

and by Proposition 5,

$$(f_1f_2f_3) = [\phi_1, [B_1^{-1}, R_2]],$$

and

$$(v_1v_2v_3)_j = (L_2F_2)^j[F_1, [R_1^{-1}F_2^{-1}R_1, B_1]](L_2F_2)^{-j} \quad (j = 0, 1).$$

A similar 3-cycle can be given for the other face orbit, and it is clear that Proposition 6 applies to all orbits. The solution can then be completed via Proposition 7.

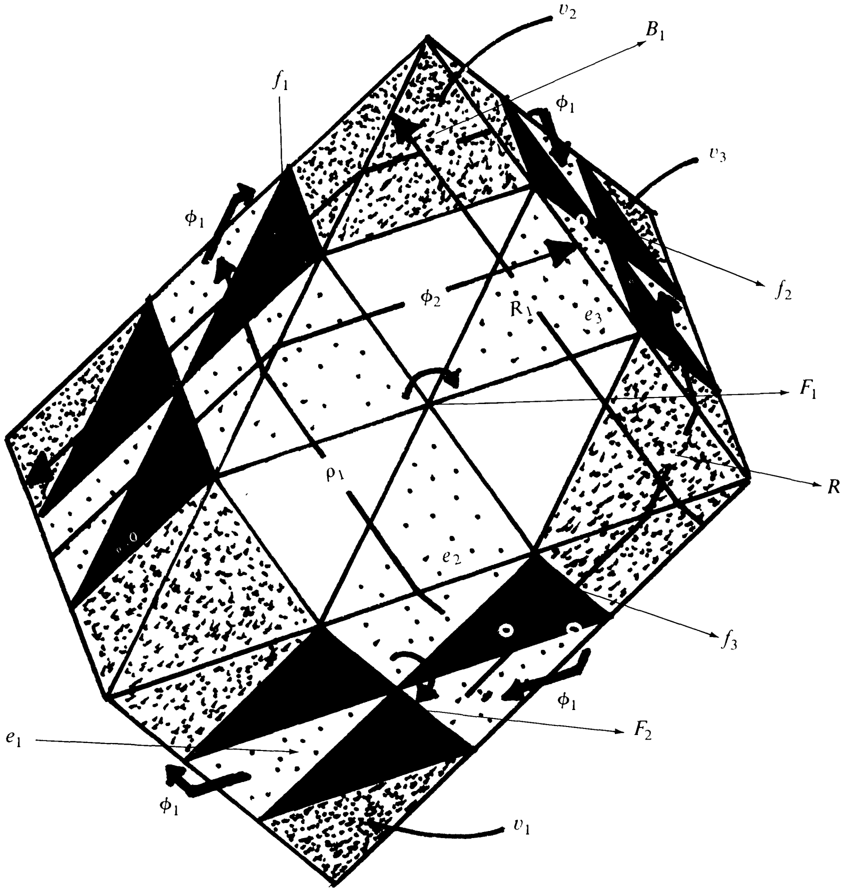


FIG. 9

Dodecahedron puzzles. Two versions will be discussed. The generators for the first one are the twelve face turns as indicated in Fig. 11(a). There are two orbits: $\mathscr{V}$ consisting of 20 corner pieces and $\mathscr{E}$ consisting of 30 edge pieces. Clearly, each generator is the product of two 5-cycles and we can proceed in the usual way, obtaining the following moves which can be used to solve the puzzle (see Fig. 11(b)):

$$(abc)_j = (R_2R_1)^j[L_1^{-1}, U^{-1}R_1U](R_2R_1)^{-j} \quad (j = 0, \pm 1)$$
$$(123)_j = (L_1L_2F)^j[R_1^{-1}, FB^{-1}U^{-1}BF^{-1}](L_1L_2F)^{-j} \quad (j = 0, 1).$$

The second puzzle, shown in Fig. 12(a), is a deeper cut version of the first one, the extreme case being the Magic Crystal of [**12**] shown in Fig. 12(b). Noting that the face pieces a and a' act as a unit (with two possible orientations), we can see that there are three orbits $\mathscr{V}$, $\mathscr{E}$, and $\mathscr{F}$ of sizes 20, 30, and 30, respectively.

The following basic 3-cycles can be obtained via Proposition 3. Oriented variations and other details are left to the reader.

$$(xyz) = [L_2^{-1}, R_1],$$
$$(abc) = [L_1^{-1}, L_2^{-1}\rho_2^2R_1\phi_2^{-2}F_1\phi_2^2R_1^{-1}\rho_2^{-2}L_2],$$
$$(123) = [L_1^{-1}, \phi_2R_2L_2^{-1}R_1^{-1}R_2^{-1}R_1\lambda_1^{-2}L_2R_2^{-1}\phi_2^{-1}].$$

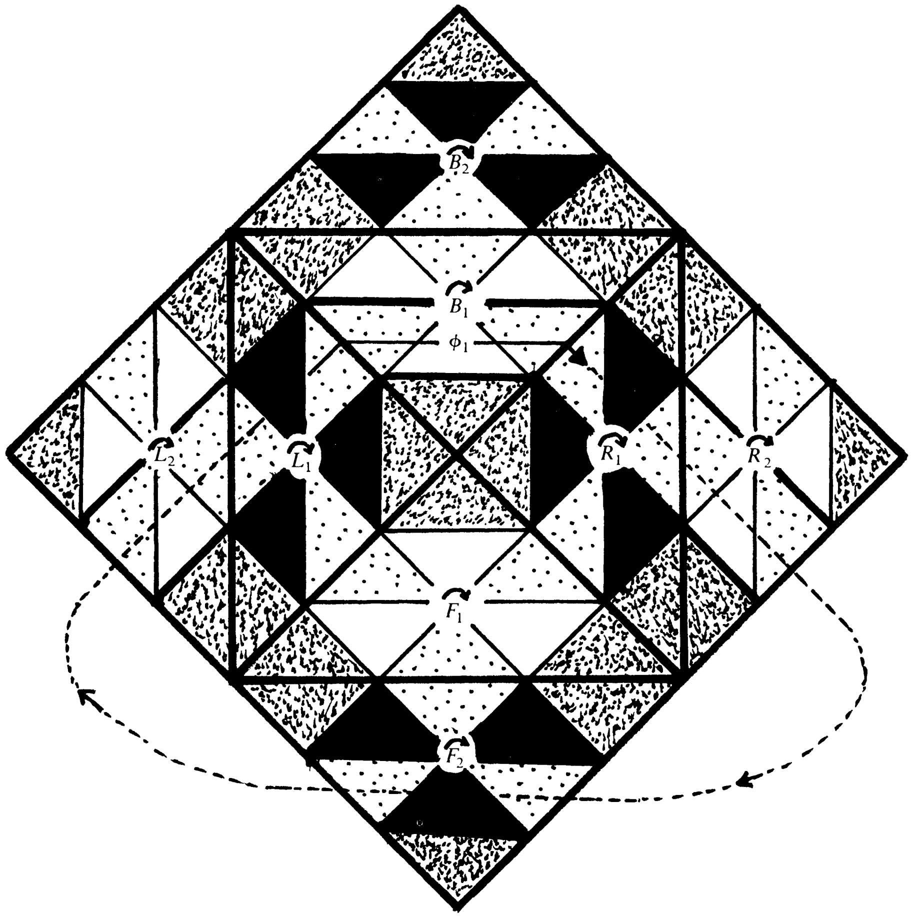
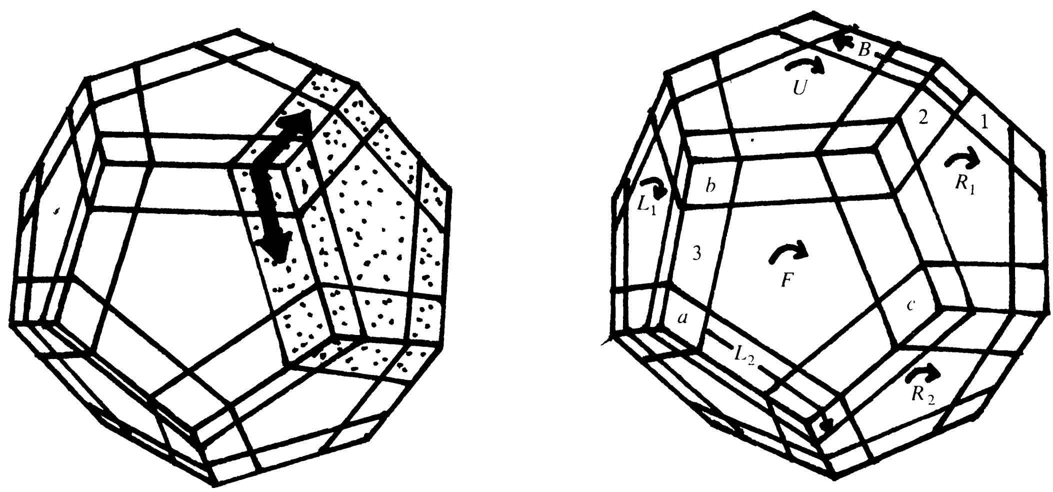
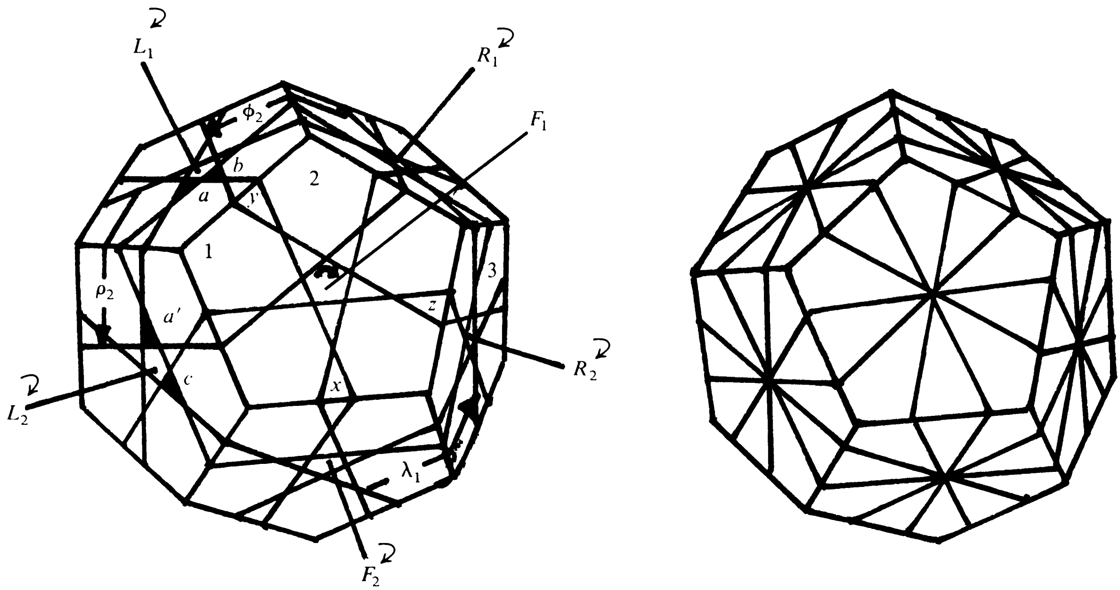


FIG. 10. Flattened octahedron puzzle with one slice turn shown.



(a) (b)

FIG. 11



(a) (b)

FIG. 12

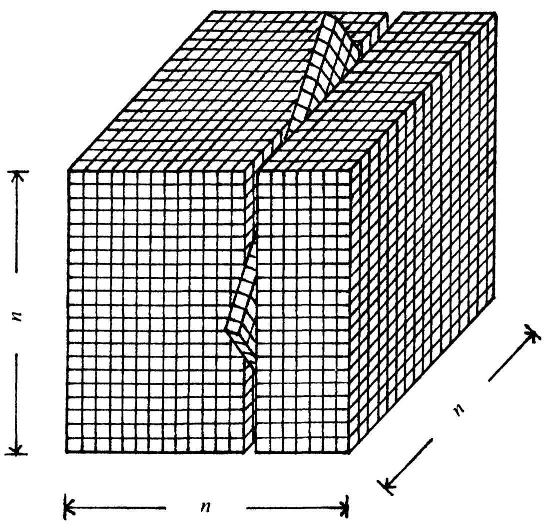


FIG. 13

The $n \times n \times n$ cube puzzle. As the ultimate cubelike puzzle, consider a cube partitioned into n^3 congruent subcubes and such that each of the $3n$ layers can be rotated as indicated in Fig. 13. The algorithm to be described is for unscrambling any generated configuration of surface cubelets. It will be seen that the algorithm leaves the $(n-2)^3$ internal cubelets invariant, given that they are fixed up in advance. Thus, if one wanted, in addition, to unscramble the internal cubelets, this could be done by applying the algorithm successively to a nested sequence of cubes starting with either a $1 \times 1 \times 1$ center or a $2 \times 2 \times 2$ center (depending on the parity of n) and working outward. Each application of the algorithm would fix up a new "shell" (the surrounding cubelets, of course, would just "come along for the ride").

We assume that n is odd: $n = 2m + 1$, $m \geqslant 2$. The even case is a corollary and is left to the reader. We set up notation in an xyz-coordinate system with the center of the cube at the origin and the edges parallel to the coordinate axes. A typical cubelet is described by the set

$$\left[i - \frac{1}{2}, i + \frac{1}{2}\right] \times \left[j - \frac{1}{2}, j + \frac{1}{2}\right] \times \left[k - \frac{1}{2}, k + \frac{1}{2}\right]$$

or more briefly by the coordinates (i, j, k) of its center. The description may seem inadequate because it ignores orientations. It is easy to see, though, that except for those cubelets corresponding to the $3 \times 3 \times 3$ case, only one orientation is possible.

Each generator is a product of $m(m+1)$ 4-cycles. Clockwise turns X_i, Y_j, Z_k $(i, j, k = 0, \pm 1, \ldots, \pm m)$ are given by:

$$X_i = \prod_{j=0}^{m} \prod_{k=1}^{m} ((i,j,k)(i,k,-j)(i,-j,-k)(i,-k,j)),$$

$$Y_j = \prod_{i=0}^{m} \prod_{k=1}^{m} ((i,j,k)(-k,j,i)(-i,j,-k)(k,j,-i)),$$

$$Z_k = \prod_{i=0}^{m} \prod_{j=1}^{m} ((i,j,k)(j,-i,k)(-i,-j,k)(-j,i,k)).$$

Here, "clockwise" is from the point of view of looking along a positive axis toward the origin.

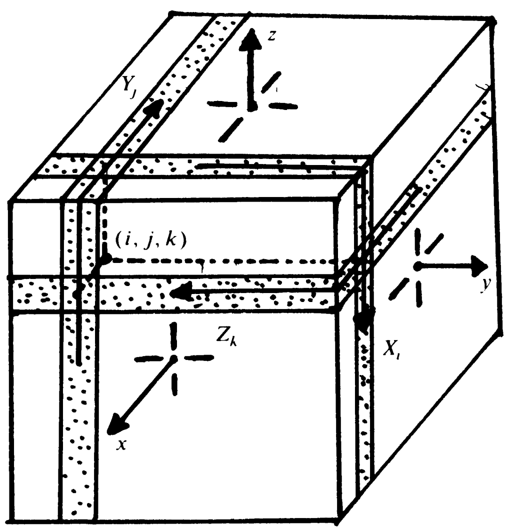


FIG. 14

See Fig. 14. Notice that the central slices X_0, Y_0, Z_0 can be omitted from the list of generators as they can be replaced by the complementary moves.

There are $m^2 + 1$ orbits (see blocked portion in Fig. 15):

$$\mathscr{V} = \mathcal{O}((m,m,m)), \quad \mathscr{E} = \mathcal{O}((m,m,0)),$$

$$\mathscr{E}_k = \mathcal{O}((m,m,k)) \quad (k = 1, \ldots, m-1),$$

and

$$\mathscr{F}_{jk} = \mathcal{O}((m,j,k)) \quad (j = 1, \ldots, m-1;\ k = 0, \pm 1, \ldots, \pm(j-1), j).$$

Orbits $\mathscr{V}$ and $\mathscr{E}$ correspond to the $3 \times 3 \times 3$ case and consist of 8 and 12 cubelets, respectively. Each $\mathscr{E}_k$ consists of 24 edge cubelets and each $\mathscr{F}_{jk}$ consists of 24 face cubelets. The shadings in Fig. 15 illustrate the four types of orbits.

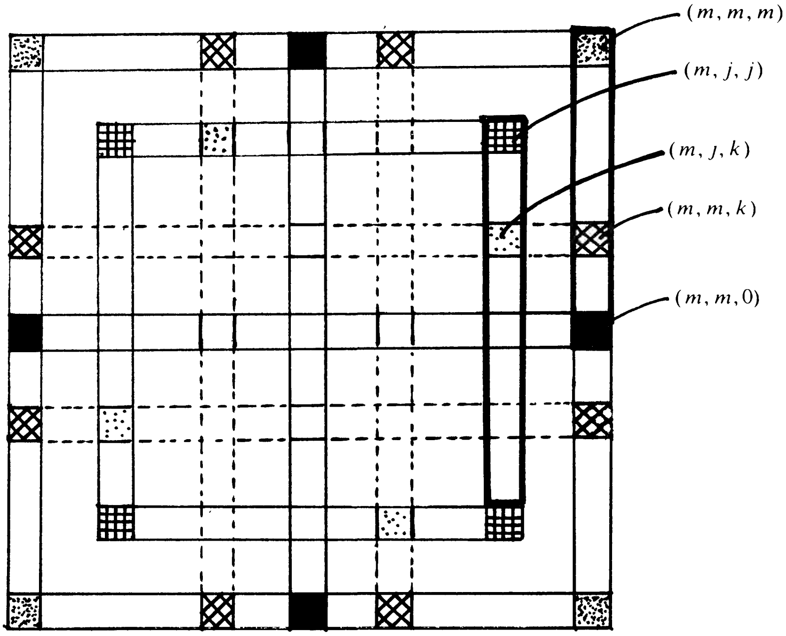


FIG. 15. Orbits in the face $x = m$.

Now consider an arbitrary generated configuration g. Let $L_1, \ldots, L_N$ be any sequence of layer turns $\neq X_0, Y_0, Z_0$ such that $g = L_1 \cdots L_N$. Let $f(\mathcal{O})$ and $s(\mathcal{O})$ denote the number of L_i's which are, respectively, face and slice turns such that $\nu(L_i, \mathcal{O}) = 1$. Then, clearly,

$$\nu(g, \mathcal{O}) = f(\mathcal{O}) + s(\mathcal{O})$$

mod 2 from which we obtain the following:

$$\begin{aligned}
\nu(g, \mathscr{F}_{jj}) &= f(\mathscr{F}_{jj}) = f(\mathscr{E}) = \nu(g, \mathscr{E}),\\
\nu(g, \mathscr{F}_{jk}) &= f(\mathscr{E}) + s(\mathscr{F}_{j0}) + s(\mathscr{F}_{0k})\\
&= \nu(g, \mathscr{E}) + \nu(g, \mathscr{F}_{j0}) + \nu(g, \mathscr{F}_{k0}) \quad \text{for } k \neq j,\\
\nu(g, \mathscr{E}_k) &= s(\mathscr{E}_k) = \nu(g, \mathscr{E}) + \nu(g, \mathscr{F}_{k0}),
\end{aligned}$$

and

$$\nu(g, \mathscr{V}) = f(\mathscr{V}) = \nu(g, \mathscr{E}).$$

Thus condition (ii) of Proposition 1 is met with $\mathcal{O}_j = \mathscr{F}_{j0}$ for $j = 1, \ldots, m-1$ and $\mathcal{O}_m = \mathscr{E}$. Furthermore, it is clear that $\nu(Y_i, \mathcal{O}_j) = \delta_{ij}$ and, therefore, condition (1) of Proposition 1 is satisfied with $h_i = Y_i$. The surprising conclusion is that all parities can be made even by making at most m quarter-turns.

Notice that if the face centers are correctly oriented at the beginning, then $\nu(g, \mathscr{E}) = 0$ and the above adjustment will not affect the face centers.

Via Propositions 3–4, we obtain the following formula which gives 3-cycles for *all* orbits simultaneously:

$$((-j, m, k)(m, j, k)(-k, m, -j)) = \left[Z_k, X_m Z_{-j} X_m^{-1}\right]$$

for $0 \leqslant |k| \leqslant j \leqslant m$, $k \neq -j$. The solution then proceeds as expected.

To see that the internal cubelets remain invariant under the algorithm if they are fixed up in

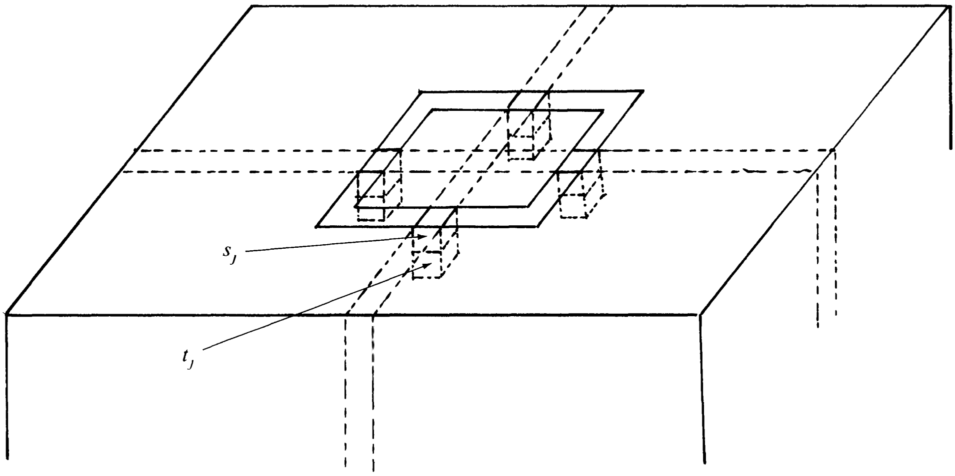


FIG. 16

advance, assume that $\nu(g,\mathcal{O}(t))=0$ for all internal cubelets t. Then, from Fig. 16, it is clear that

$$\nu(g,\mathcal{O}_j)=\nu\big(g,\mathcal{O}(s_j)\big)=\nu\big(g,\mathcal{O}(t_j)\big)=0 \quad \text{for} \quad j=1,\dots,m-1.$$

Hence, in this situation, parity adjustments which would affect the internal cubelets are not needed. Therefore, since all other moves are commutators, the internal cubelets are unaltered by the algorithm.

6. The underlying groups. If, in the above examples, we ignore the matter of orienting pivotal centers, the underlying group P for such a puzzle can be described as a subdirect product of groups of the form $H_i \wr G_i$, where $H_i=\mathbb{Z}_{r_i}$ and G_i is either $\operatorname{Alt}\mathcal{X}_i$ or $\operatorname{Sym}\mathcal{X}_i$. By Propositions 2 and 7, $P|H_i \wr G_i$ has index r_i in $H_i \wr G_i$. Also, if $\{\mathcal{X}_1,\dots,\mathcal{X}_m\}$ is a parity basis in the sense of Proposition 1, then in forming a typical element $g=g_1\cdots g_m\cdots g_N$, one has $|\mathcal{X}_i|!$ choices for each g_i, $i=1,\dots,m$, and $\frac{1}{2}|\mathcal{X}_i|!$ choices for each of the remaining g_i's. It follows that the order of P is equal to

$$\frac{1}{2^{N-m}}\prod_{i=1}^{N}|\mathcal{X}_i|!\,r_i^{|\mathcal{X}_i|-1}.$$

In particular, we see that the $(2m+1)\times(2m+1)\times(2m+1)$ cube contains

$$\frac{8!\cdot 12!\cdot 24!^{m^2-1}\cdot 2^{11}\cdot 3^7}{2^{m^2+1-m}}$$

distinct patterns. For the $(2m)\times(2m)\times(2m)$ cube, the reader will find that there are m^2+1-m orbits, one of size 8 and all others of size 24, and that a parity basis has size m. Hence, taking into account the fact that the 24 rigid motions of the cube yield duplications, the $(2m)\times(2m)\times(2m)$ cube has

$$\frac{8!\cdot 24!^{m(m-1)}\cdot 3^7}{2^{(m-1)^2}\cdot 24}$$

distinct patterns. In case the faces are solidly colored, the number of distinguishable patterns decreases. The following table summarizes our findings for the cases $n=2,3,4$ and 5. Conflicting values for the case $n=4$ have appeared in print.

Colored cube size	Number of distinguishable patterns
$2 \times 2 \times 2$	$\frac{8! \cdot 3^7}{24}$
$3 \times 3 \times 3$	$\frac{8! \cdot 12! \cdot 2^{11} \cdot 3^7}{2}$
$4 \times 4 \times 4$	$\frac{8! \cdot 24!^2 \cdot 3^7}{2 \cdot 24 \cdot (4!^6/2)}$
$5 \times 5 \times 5$	$\frac{8! \cdot 12! \cdot 24!^3 \cdot 2^{11} \cdot 3^7}{2^3 \cdot (4!^6/2)^2}$

7. Efficiency. Many of the algorithms outlined in §5 can be shortened considerably by replacing 3-cycles by *relative* 3-cycles. If $\mathcal{O}_1, \ldots, \mathcal{O}_N$ is any ordering of the orbits, then the "pure" 3-cycles used to unscramble the orbit $\mathcal{O}_i$ can be replaced by 3-cycles relative to $\mathcal{O}_1 \cup \cdots \cup \mathcal{O}_i$. These, in general, will have shorter factorizations, and can still be obtained via Proposition 3. A measurement of the efficiency of any particular ordering is given by

$$(*) \qquad m(\mathcal{O}_1, \ldots, \mathcal{O}_N) = \sum_{i=1}^{N} l_i |\mathcal{O}_i|,$$

where l_i denotes the shortest possible 3-cycle in $\mathcal{O}_i$ relative to $\mathcal{O}_1 \cup \cdots \cup \mathcal{O}_i$.

The following table is a summary of our findings of most efficient orderings with the relevant additional 3-cycles given in the right-hand column (cf. Figs. 4, 6, 9, 11, 12, 14). The findings are based on minimizing $(*)$ over the $N!$ orderings with the l_i's taken as lengths of commutator 3-cycles.

Puzzle	Most efficient ordering	Relative 3-cycles
$3 \times 3 \times 3$ cube	$(\mathscr{E}, \mathscr{V})$	$[R, F]$
tetrahedron #2	$(\mathscr{F}_1, \mathscr{F}_2, \mathscr{F}_3, \mathscr{V})$	$[\overline{RT}, \overline{BT}], [\overline{BT}, \overline{LT}], [\overline{LT}, \overline{RT}]$
octahedron #2	$(\mathscr{E}, \mathscr{V}, \mathscr{F}_1, \mathscr{F}_2)$	$[F_1, F_2], [F_1, B_1]$
dodecahedron #1	$(\mathscr{E}, \mathscr{V})$	$[R_1, R_2]$
dodecahedron #2	$(\mathscr{F}, \mathscr{V}, \mathscr{E})$	$[F_1, F_2], [L_1, R_1^{-1} R_2^{-1} R_1]$
$n \times n \times n$ cube	$(\mathcal{O}(m,1,0), \ldots, \mathcal{O}(m,m,0)$, then any order)	$[X_m, Y_1], \ldots, [X_m, Y_m]$

References

1. Christoph Bandelow, Inside Rubik's Cube and Beyond, Birkhäuser, Boston, 1982.
2. E. R. Berlekamp, J. H. Conway, and R. K. Guy, Winning Ways for Your Mathematical Plays, Vol. 2, Academic Press, New York, 1982.
3. Garrett Birkhoff and Saunders Mac Lane, A Survey of Modern Algebra, Macmillan, New York, 1941.
4. F. J. Budden, The Fascination of Groups, Cambridge University Press, London, 1972.
5. W. Burnside, Theory of Groups of Finite Order, Cambridge University Press, London, 1897.
6. Robert Carmichael, Introduction to the Theory of Groups of Finite Order, Ginn, Boston, 1937.
7. A. H. Frey and D. Singmaster, Handbook of Cubik Math, Enslow Publishers, Hillside, NJ, 1982.
8. Solomon W. Golomb, Rubik's cube and a model of quark confinement, Amer. J. Phys., 49 (11), November 1981.
9. Marshall Hall, The Theory of Groups, 2nd ed., Chelsea, New York, 1976.
10. I. N. Herstein, Topics in Algebra, Blaisdell, New York, 1964.

11. Douglas R. Hofstadter, Metamagical Themas: The Magic Cube's cubies are twiddled by cubists and solved by cubemeisters, Scientific American, March 1981, pp. 20–39.

12. ______, Metamagical Themas: Beyond Rubik's Cube: Spheres, pyramids, dodecahedrons and God knows what else, Scientific American, July 1982, pp. 16–31.

13. M. I. Kargapolov and Ju. I. Merzljakov, Fundamentals of the Theory of Groups, 2nd ed., Springer-Verlag Graduate Texts in Mathematics, New York, 1979.

14. Derek J. S. Robinson, A Course in the Theory of Groups, Springer-Verlag, New York, 1982.

15. John S. Rose, A Course on Group Theory, Cambridge University Press, London, 1978.

16. Joseph J. Rotman, The Theory of Groups: An Introduction, Allyn and Bacon, Boston, 1968.

17. Chris Rowley, The group of the Hungarian Magic Cube, Algebraic Structures and Applications, Proceedings of the First Western Australian Conference on Algebra, 1982.

18. Eugene Schenkman, Group Theory, Van Nostrand, Princeton, NJ, 1965.

19. David Singmaster, Notes on Rubik's Magic Cube, Enslow Publishers, Hillside, NJ, 1981.

20. John C. Thompson, Rational functions associated to presentations of finite groups, J. Algebra, 71 (1981) 481–489.

21. Helmut W. Wielandt, Finite Permutation Groups, Academic Press, New York, 1964.

22. ______, Permutation Groups through Invariant Relations and Invariant Functions, Lecture Notes, Ohio State University, 1969.

23. Richard M. Wilson, Graph puzzles, homotopy and the alternating group, J. Combin. Theory Ser. B, 16 (1974) 86–96.

AN ALTERNATIVE APPROACH TO CANONICAL FORMS OF MATRICES

E. J. P. GEORG SCHMIDT*

Department of Mathematics, McGill University, Montreal, Quebec, Canada, H3A 2K6

0. Introduction. The purpose of this paper is to describe a new approach to the derivation of the main structure theorems for linear transformations and matrices. The approach is unconventional in that it reverses the traditional order between the study of various canonical forms and the structure of exponentials of matrices (as treated, for example, in the beautiful book [**2**]). It is efficient in that it leads in a unified and direct way to the main facts needed for the derivation of canonical forms.

This approach is accessible to students who have mastered the topics in linear algebra which usually precede the structure theorems (as given, for example, in Chapters I to IX of [**3**]). It is particularly appropriate within the context of a differential equations course, where it is often necessary to review or supplement a previous treatment of linear algebra. The exposition is given for real and complex matrices, but can easily be adapted to a more abstract treatment.

1. The structure of exp[At]. Let $A = [A_{ij}]$ be an $n \times n$ real or complex matrix and let the totality of such matrices be denoted by $M_n(R)$ or $M_n(C)$, respectively. It is customary in the treatment of linear systems of differential equations to define

$$\exp[At] = \sum_{k=0}^{\infty} \frac{A^k t^k}{k!}.$$

E. J. P. Georg Schmidt: I was born and raised in South Africa, where I completed my undergraduate training. I received my Ph.D. from Stanford University working under the supervision of Ralph Phillips. After a year at the M.R.C. in Madison and two years in Aarhus, Denmark, I moved to Canada and McGill University. My mathematical interests are centered around partial differential equations. Otherwise I enjoy camping and hiking with my family, and play tennis less frequently and less well than I would like to.

*This work was partially supported by a Grant of the Natural Sciences and Engineering Research Council of Canada.

The convergence of this series is easily proved by comparison with a convergent scalar series using the norm

$$|A|_\infty = \max\{|A_{ij}|: 1 \leqslant i, j \leqslant n\},$$

as well as the estimate $|A^k|_\infty \leqslant n^{k-1}|A|_\infty^k$.

A representation of $\exp[At]$ can be obtained by examining carefully a well-known method of computing it, which can be found, for example, in [**1**]. This method exploits the Cayley-Hamilton theorem, which states that if $A \in M_n(C)$, then $p_A(A) = 0$, where $p_A(\lambda) = \det[\lambda I - A]$ is the characteristic polynomial of A. (A simple proof of this can be found in [**3**], page 211.)

We shall say that f: $C \to C$ is entire if it has a power series expansion $f(\lambda) = \sum_{k=0}^{\infty} a_k \lambda^k$ convergent for all $\lambda \in C$. Using the norm $|A|_\infty$, we can easily check that $f(A) = \sum_{k=0}^{\infty} a_k A^k$ is convergent.

We begin with two essential lemmas.

1.1. LEMMA. (a) *Let $f(\lambda)$ be entire and suppose that $p(\lambda)$ is a polynomial of degree n. Then there exists a unique entire function $g(\lambda)$ and a unique polynomial $q(\lambda)$ of degree less than n, such that $f(\lambda) = g(\lambda)p(\lambda) + q(\lambda)$.*

(b) *If $p(\lambda) = c\prod_{i=1}^{r}(\lambda - \lambda_i)^{n_i}$ (in which case $\sum_{i=1}^{r} n_i = n$), then $q(\lambda) = \sum_{k=0}^{n-1} q_k \lambda^k$ is uniquely determined by the following system of n linear equations in the n unknown coefficients q_k:*

$$D^j q(\lambda_i) = D^j f(\lambda_i), \quad i = 1 \text{ to } r, \quad j = 0 \text{ to } n_i - 1,$$

(where $D = \dfrac{d}{d\lambda}$).

The proof—an easy exercise for the instructor—uses induction on n, as well as basic facts concerning polynomials and the rearrangement of power series.

This lemma can be used to evaluate $f(A)$ for any entire f. One simply expresses f in the form

$$f(\lambda) = g(\lambda)p_A(\lambda) + q(\lambda)$$

and then finds, since $p_A(A) = 0$, that $f(A) = q(A)$. Since the lemma tells one how to compute q, this is an effective computational procedure.

1.2. LEMMA. *Let $\{Q_k(t)\}_{k=1}^{m}$ be matrix polynomials, i.e., polynomials with matrix coefficients, and $\{\mu_k\}_{k=1}^{m}$ be distinct numbers. Then, if $\sum_{k=1}^{m} e^{\mu_k t}Q_k(t) = 0$ for all t, each $Q_k(t)$ is identically zero.*

This is proved by induction on m. For $m = 1$ it is trivial. Given a relation between m terms, one multiplies through by $e^{-\mu_m t}$ and differentiates $1 + \deg(Q_m)$ times to obtain a relation in $m - 1$ terms involving polynomials $\{\tilde{Q}_k(t)\}_{k=1}^{m-1}$ with $\deg(\tilde{Q}_k) = \deg(Q_k)$. By the induction hypothesis $\tilde{Q}_k$, and hence also Q_k, vanishes identically.

1.3. THEOREM. *Let $A \in M_n(C)$ have characteristic polynomial $p_A(\lambda) = \prod_{i=1}^{r}(\lambda - \lambda_i)^{n_i}$. Then there exists a unique decomposition*

$$\exp[At] = \sum_{i=1}^{r} e^{\lambda_i t}\left[P_i + \sum_{k=1}^{k_i - 1} N_i^k \frac{t^k}{k!}\right],$$

where the sum over k is understood to be 0 when $k_i = 1$, and where

(a) *P_i, N_i are commuting complex matrices each of which can be obtained by evaluating a suitable complex polynomial in A;*

(b) *$P_i P_j = \delta_{ij} P_i$ (with $\delta_{ij} = 1$ if $i = j$, 0 otherwise) and $\sum_{i=1}^{r} P_i = I$ (the identity matrix); thus*

each P_i *is a projection onto a subspace* V_i *of* C^n *and* C^n *is the direct sum of these subspaces*;

(c) N_i *is nilpotent of order* $k_i \leqslant n_i$ (*so* $N_i^{k_i} = 0$ *and, if* $k_i > 1$, $N_i^{k_i - 1} \neq 0$), $P_j N_i = N_i P_j = \delta_{ij} N_i$ (*so that* $N_i(V_i) \subset V_i$ *and* $N_i(V_j) = \{0\}$ *for* $j \neq i$);

(d) $A = \sum_{i=1}^{r}(P_i + \lambda_i N_i)$;

(e) V_i *has dimension* n_i;

(f) *the minimal polynomial of* A *is* $p_A^{\min}(\lambda) = \prod_{i=1}^{r} (\lambda - \lambda_i)^{k_i}$.

Proof. We apply Lemma 1.1 with $f(\lambda) = \exp[t\lambda]$ and $p(\lambda) = p_A(\lambda)$. Then

$$\exp[tA] = \sum_{k=0}^{n-1} q_k(t) A^k,$$

where the n coefficients $q_k(t)$ are *uniquely* determined for each t, as solutions of the linear system of n equations

$$\sum_{k=j}^{n-1} \left[\frac{k!}{(k-j)!}\lambda_i^{k-j}\right] q_k(t) = t^j e^{\lambda_i t}$$

corresponding to $i = 1$ to r, and $j = 0$ to $n_i - 1$. The coefficient matrix of this system is independent of t and invertible (invertibility being equivalent to uniqueness); thus the solutions have the form

$$q_k(t) = \sum_{i=1}^{r} \sum_{j=0}^{n_i - 1} q_{ijk} t^j e^{\lambda_i t}.$$

Hence

$$(*) \qquad \exp[tA] = \sum_{k=0}^{n-1} q_k(t) A^k = \sum_{i=1}^{r} e^{\lambda_i t} Q_i(t),$$

where

$$Q_i(t) = \sum_{j=0}^{n_i - 1} Q_{ij} \frac{t^j}{j!} \quad \text{with} \quad Q_{ij} = j! \sum_{k=0}^{n-1} q_{ijk} A^k.$$

We note that $Q_i(t)$ may have degree less than $n_i - 1$ and define $k_i = \deg(Q_i) + 1$.

The special structure of the polynomials $Q_i(t)$, which is the main content of the theorem, is a consequence of the identity $\exp[A(s+t)] = \exp[As]\exp[At]$, which can readily be proved using the power series definition. This implies

$$\sum_{i=1}^{r} e^{\lambda_i (s+t)} Q_i(s+t) = \left[\sum_{i=1}^{r} e^{\lambda_i s} Q_i(s)\right]\left[\sum_{j=1}^{r} e^{\lambda_j t} Q_j(t)\right],$$

or,

$$\sum_{i=1}^{r} e^{\lambda_i s}\left[e^{\lambda_i t} Q_i(s+t) - Q_i(s)\sum_{j=1}^{r} e^{\lambda_j t} Q_j(t)\right] = 0.$$

Fix t for the moment; it follows from Lemma 1.2 that the square bracketed expression vanishes for each i; rewriting this, we have for $i = 1$ to r that

$$\sum_{j=1}^{r} e^{\lambda_j t}\left[Q_i(s) Q_j(t) - \delta_{ij} Q_i(s+t)\right] = 0.$$

Again by Lemma 1.2 it follows that

$$(**) \qquad Q_i(s) Q_j(t) = \delta_{ij} Q_i(s+t).$$

We now define $P_i = Q_{i0}$ and $N_i = Q_{i1}$ (or 0 if $n_i = 1$); (a) follows. Noting that $Q_{i0} = Q_i(0)$, one sets $t = 0$ in $(*)$ to find $\sum_{i=1}^r P_i = I$, and $s = t = 0$ in $(**)$ to obtain $P_i P_j = \delta_{ij} P_i$. Property (b) follows from these two facts.

Noting that $Q_{ij} = D^j Q_i(0)$ (with D denoting differentiation w.r.t. t), we differentiate $(**)$ once with respect to s and set $s = t = 0$ to find $N_i P_j = \delta_{ij} N_i$. Differentiating $(**)$ l times with respect to s and k times with respect to t and then setting $s = t = 0$ and $i = j$, one finds $Q_{il} Q_{ik} = Q_{i(l+k)}$ if $l + k < k_i$, and $Q_{il} Q_{ik} = 0$ if $l + k \geqslant k_i$. Hence $Q_{ij} = Q_{i1}^j = N_i^j$ for $j < k_i$ and $N_i^{k_i} = 0$. This completes the proof of (c), as well as of the representation of $\exp[At]$. Differentiating the latter with respect to t and setting $t = 0$, one obtains the representation (d) for A. The uniqueness of the representation of $\exp[At]$ is proved by applying Lemma 1.2 to the difference of two such representations.

To prove (e) note first that $A(V_i) \subset V_i$. Let A_i be the restriction of A to V_i. It is then easily seen, since N_i has only 0 as an eigenvalue and since P_i acts as the identity on V_i, that

$$p_{A_i}(\lambda) = (\lambda - \lambda_i)^{\dim V_i}.$$

Therefore

$$p_A(\lambda) = \prod_{i=1}^r p_{A_i}(\lambda) = \prod_{i=1}^r (\lambda - \lambda_i)^{\dim V_i}$$

from which (e) follows. Finally (f) is left to the reader as an exercise.

We note an immediate corollary of the proof of the last theorem.

1.4. COROLLARY. *If $A \in M_n(R)$ and its eigenvalues are all real, then in Theorem* 1.3 *the matrices P_i and N_i are real and the V_i are subspaces of R^n.*

Theorem 1.3 and Corollary 1.4 contain essential ingredients for proving a variety of structure theorems and for the derivation of canonical forms. We illustrate some of these below.

2. Normal matrices. A matrix $A \in M_n(C)$ is said to be *normal* if $A^*A = AA^*$, where A^* is the "adjoint" of A defined to be the complex conjugate of the transposed matrix (i.e., $A^* = \overline{A^t}$). It is natural to consider in C^n the inner product

$$\langle z, w \rangle = z \cdot \overline{w} = \sum_{k=1}^n z_k \overline{w}_k,$$

and the associated norm $|z| = \langle z, z \rangle^{1/2}$. It is easy to verify that for any $A \in M_n(C)$ and any $z, w \in C^n$, $\langle Az, w \rangle = \langle z, A^* w \rangle$ and $(A^*)^* = A$.

The main elementary facts concerning normal matrices are given in the following readily proved lemma.

2.1. LEMMA. (a) *Let $A \in M_n(C)$ be normal. Then $|Az| = |A^*z|$.*

(b) *Let $P \in M_n(C)$ be a normal projection. Then $P^* = P$ and P projects orthogonally onto the subspace $V = P(C^n)$ of C^n.*

(c) *Let $N \in M_n(C)$ be nilpotent and normal. Then $N = 0$.*

This, together with Theorem 1.3, directly yields the spectral theorem and the diagonalisation theorem for normal matrices.

2.2. THEOREM. *Let $A \in M_n(C)$ be normal. Then*

(a) $A = \sum_{i=1}^r \lambda_i P_i$, *where P_i is the orthogonal projection onto the n_i dimensional subspace $V_i = \{z \in C^n : Az = \lambda_i z\}$; these subspaces are orthogonal and span C^n. The minimal polynomial is* $\prod_{i=1}^r (\lambda - \lambda_i)$.

(b) *One can find a unitary matrix* $U \in M_n(C)$ *(i.e.,* $U^*U = UU^* = I$*) such that* $U^*AU = D$*, where* D *is the diagonal matrix successively having* n_i λ_i*'s along the diagonal.*

To prove (a) one simply notes that the matrices P_i and N_i are normal (since they are obtained as polynomials in A) and then applies the lemma. To show (b) one defines U to be a matrix having as its columns successive orthonormal basis vectors for the subspaces V_i.

3. Symmetric and real normal matrices. Consider matrices $A \in M_n(R)$ such that $A^tA = AA^t$. This class of matrices includes, in particular, symmetric matrices (for which $A^t = A$), skew-symmetric matrices (for which $A^t = -A$) and orthogonal matrices (for which $O^tO = OO^t = I$).

Symmetric matrices have real eigenvalues. Hence we can apply Corollary 1.4 and parallel the proof of Theorem 2.2 (with R replacing C) to obtain the following result.

3.1. THEOREM. *Let* $A \in M_n(R)$ *be symmetric. Then all its eigenvalues are real and*

(a) $A = \sum_{i=1}^{r} \lambda_i P_i$*, where* P_i *is the orthogonal projection onto the* n_i *dimensional subspace* $V_i = \{x \in R^n\colon Ax = \lambda_i x\}$*; these subspaces are orthogonal and span* R^n.

(b) *One can find an orthogonal matrix* $O \in M_n(R)$ *such that* $O^tAO = D$*, where* D *is the diagonal matrix successively having* n_i λ_i*'s along the diagonal.*

For other real normal matrices the eigenvalues will not all be real. It is then of interest to find a structure theorem or canonical form within $M_n(R)$. These are both provided by the next theorem.

3.2. THEOREM. *Let* $A \in M_n(R)$ *be normal. The complex eigenvalues come in conjugate pairs having the same multiplicities; we suppose these are*

$$\lambda_{2j-1} = \mu_j + i\nu_j, \quad \lambda_{2j} = \mu_j - i\nu_j$$

for $j = 1$ *to* m *(each having multiplicity* n_j^c*), and that the eigenvalues* λ_i *for* $i = 2m+1$ *to* r *are real. Then*

(a) $A = \sum_{j=1}^{m} [\mu_j P_j^c - \nu_j R_j^c] + \sum_{i=2m+1}^{r} \lambda_i P_i$*, where*

(i) P_j^c*,* R_j^c*,* P_i *are real matrices commuting with each other;*

(ii) *For* $j = 1$ *to* m*,* P_j^c *is the orthogonal projection onto a subspace* V_j^c *of* R^n *which has dimension* $2n_j^c$ *and such that* $A\colon V_j^c \to V_j^c$ *has eigenvalues* $\mu_j \pm i\nu_j$ *each with multiplicity* n_j^c*;*

(iii) *For* $i = 2m+1$ *to* r*,* P_i *is the orthogonal projection onto the eigenspace* $V_i = \{x \in R^n\colon Ax = \lambda_i x\}$*;*

(iv) R^n *is the orthogonal direct sum of the subspaces* V_j^c *(*$j = 1$ *to* m*) and* V_i *(*$i = 2m+1$ *to* r*);*

(v) $[R_j^c]^2 = -P_j^c$ *and* $P_j^cR_j^c = R_j^c$ *(so that* $R_j^c(V_j^c) \subset V_j^c$ *and* R_j^c *annihilates the other subspaces).*

(b) *One can find an orthogonal matrix* $O \in M_n(R)$ *such that* $O^tAO = C$*, where* C *is a tridiagonal matrix with* 2×2 *matrices* $\begin{bmatrix} \mu_j & \nu_j \\ -\nu_j & \mu_j \end{bmatrix}$ *repeated* n_j^c *times along the diagonal, and* λ_i *repeated* n_i *times.*

Proof. Considering A as a matrix in $M_n(C)$, one can apply Theorem 1.3 and Lemma 2.1(c) to find

$$\exp[At] = \sum_{j=1}^{m} \left[e^{(\mu_j + i\nu_j)t} P_{2j-1} + e^{(\mu_j - i\nu_j)t} P_{2j} \right] + \sum_{i=2m+1}^{r} e^{\lambda_i t} P_i.$$

Conjugating this, noting that $\exp[At] = \exp[\bar{A}t]$ and using Lemma 1.2, one finds that $P_{2j} = \bar{P}_{2j-1}$ for $j = 1$ to m, and that P_i is real for $i = 2m+1$ to r. Now we define

$$P_j^c = 2\,\mathrm{Re}(P_{2j-1}) = P_{2j-1} + P_{2j}$$

and

$$R_j^c = 2\,\mathrm{Im}(P_{2j-1}) = -i(P_{2j-1} - P_{2j}),$$

and then obtain the stated representation for A by multiplying out the complex terms in

$$A = \sum_{i=1}^{m}\left[(\mu_j + i\nu_j)\frac{1}{2}(P_j^c + iR_j^c) + (\mu_j - i\nu_j)\frac{1}{2}(P_j^c - iR_j^c)\right] + \sum_{i=2m+1}^{r} \lambda_i P_i.$$

Most of the properties of P_j^c and R_j^c follow immediately from Lemma 2.1(b), and from the identities

$$P_{2j-1}^2 = P_{2j-1} \quad \text{and} \quad P_{2j-1}\overline{P}_{2j-1} = P_{2j-1}P_{2j} = 0.$$

Only (ii) and (iv) need further attention.

To prove (iv) note first that, since $P_iP_j = 0$ (for $i \neq j$ and $1 \leqslant i,\ j \leqslant r$), it is also true that $P_j^cP_k^c = 0$ (for $j \neq k$ and $1 \leqslant j,\ k \leqslant m$) and $P_j^cP_i = 0$ (for $1 \leqslant j \leqslant m$ and $2m+1 \leqslant i \leqslant r$). Hence the subspaces $V_j^c = P_j^c(R^n)$ (for $j = 1$ to m) and $V_i = P_i(R^n)$ (for $i = 2m+1$ to r) are mutually orthogonal. Since

$$\sum_{j=1}^{m} 2n_j^c + \sum_{i=2m+1}^{r} n_i = n,$$

(iv) follows if one can prove that $\dim V_i = n_i$ and $\dim V_j^c = 2n_j^c$, where the dimension is to be taken over the reals. To show that $\dim V_i = n_i$ note that V_i is just the space spanned by the columns of P_i (which are real); the real and complex dimensions of this column space are the same and the latter is known to be n_i. To prove $\dim V_j^c = n_j^c$ (and to prove the remaining assertion of (ii)) one chooses an orthonormal basis $\{x^k + iy^k\}_{k=1}^{n_j^c}$ for $V_{2j-1} = P_{2j-1}(C^n)$; then $\{x^k - iy^k\}_{k=1}^{n_j^c}$ is an orthonormal basis of $V_{2j} = \overline{V}_{2j-1}$. It then follows easily from the identities

$$\langle x^l + iy^l, x^k + iy^k\rangle = \delta_{lk} \quad \text{and} \quad \langle x^l + iy^l, x^k - iy^k\rangle = 0$$

that $B_j = \{x^1, y^1, x^2, y^2, \ldots, x^{n_j}, y^{n_j}\}$ is an orthonormal basis of real vectors for $V_{2j-1} + V_{2j}$ (a subspace of C^n). If now $y \in V_j^c$ then $y = (P_{2j-1} + P_{2j})x$, so that $y \in V_{2j-1} + V_{2j}$ and hence can be expressed as a (real) linear combination of the vectors in B_j; thus B_j is a basis for V_j^c which therefore has dimension $2n_j^c$.

To complete the proof of (ii) it is necessary to show that the restriction A_j of A to V_j^c has eigenvalues $\mu_j \pm i\nu_j$. It follows from

$$A(x^k + iy^k) = (\mu_j + i\nu_j)(x^k + iy^k)$$

that $A_jx^k = \mu_jx^k - \nu_jy^k$ and $A_jy^k = \nu_jx^k + \mu_jy^k$, so the matrix of A_j with respect to the basis B_j is a tridiagonal $2n_j \times 2n_j$ matrix having 2×2 blocks $\begin{bmatrix} \mu_j & \nu_j \\ -\nu_j & \mu_j \end{bmatrix}$ along the diagonal; this matrix is readily seen to have eigenvalues $\mu_j \pm i\nu_j$.

Now (b) is also immediate. Along with the already constructed orthonormal bases for V_j^c ($j = 1$ to m) one chooses also orthonormal bases for V_i ($i = 2m+1$ to r) and lets O be the matrix having the successive vectors of these bases as columns.

We state the next result because of its relevance to differential equations.

3.3. COROLLARY. *Let $A \in M_n(R)$ be normal. Then in the notation of the last theorem*

$$\exp[At] = \sum_{j=1}^{m} e^{\mu_j t}\left[\cos(\nu_j t)P_j^c - \sin(\nu_j t)R_j^c\right] + \sum_{i=2m+1}^{r} e^{\lambda_i t}P_i.$$

Moreover $\cos(\nu_j t)P_j^c - \sin(\nu_j t)R_j^c = O_j(t)$ *is orthogonal and satisfies* $O_j(s+t) = O_j(s)O_j(t)$.

4. A structure theorem for arbitrary $A \in M_n(R)$. If $A \in M_n(R)$ has some complex eigenvalues, then Theorem 1.3 involves complex matrices. We indicate how to derive a structure theorem for A involving only real matrices. As in Section 3 suppose that the complex eigenvalues of A are $\lambda_{2j-1} = \mu_j + i\nu_j$, $\lambda_{2j} = \mu_j - i\nu_j$ (each with multiplicity n_j^c) for $j = 1$ to m and that λ_i is real for $i = 2m+1$ to r. As in the proof of Theorem 3.1, one has that $P_{2j} = \overline{P}_{2j-1}$ and $N_{2j} = \overline{N}_{2j-1}$ for

$j = 1$ to m, while P_i and N_i are real for $i = 2m + 1$ to r. Then, defining

$$P_j^c = 2\operatorname{Re} P_{2j-1}, \quad R_j^c = 2\operatorname{Im} P_{2j-1}, \quad N_j^c = 2\operatorname{Re} N_{2j-1}$$

one has the following result.

4.1. THEOREM. $A = \sum_{j=1}^{m} (\mu_j P_j^c + \nu_j R_j^c + N_j^c) + \sum_{i=2m+1}^{r} (\lambda_i P_i + N_i)$, *where*

(i) $P_j^c, R_j^c, N_j^c, P_i, N_i$ *are real and commute*;

(ii) P_j^c $(j = 1$ *to* $m)$ *is a projection onto a subspace* V_j^c *of* R^n *which has dimension* $2n_j^c$ *and such that the restriction of* A *to* V_j^c *has eigenvalues* $\mu_j \pm i\nu_j$ *each with multiplicity* n_j^c;

(iii) P_i $(i = 2m + 1$ *to* $r)$ *is a projection onto* $V_i = \{x \in R^n\colon (A - \lambda_i)^{k_i} = 0\}$;

(iv) R^n *is the direct sum of the subspaces* V_j^c $(j = 1$ *to* $m)$ *and* V_i $(i = 2m + 1$ *to* $r)$; P_j^c *annihilates all the subspaces except* V_j^c *and* P_i *annihilates all but* V_i;

(v) $(R_j^c)^2 = -P_j^c$ *and* $P_j^c R_j^c = R_j^c$;

(vi) N_j^c *is nilpotent of order* $k_j^c = k_{2j} = k_{2j-1}$ *and* $P_j^c N_j^c = N_j^c$.

The easy proof is similar to that of Theorem 2.1(a). That N_j^c is nilpotent of order k_j^c follows from $N_{2j-1}N_{2j} = 0$, which implies that $(N_j^c)^k = N_{2j-1}^k + N_{2j}^k$.

One can also readily prove

4.2. COROLLARY. *Let* $A \in M_n(R)$. *In the notation of the last theorem*,

$$\exp[At] = \sum_{j=1}^{m} e^{\mu_j t}\left[\cos(\nu_j t) P_j^c - \sin(\nu_j t) R_j^c\right]\left[P_j^c + \sum_{k=1}^{k_j^c - 1} \left(N_j^c\right)^k \frac{t^k}{k!}\right]$$
$$+ \sum_{i=2m+1}^{r} e^{\lambda_i t}\left[P_i + \sum_{k=1}^{k_i - 1} N_i^k \frac{t^k}{k!}\right].$$

Here $\cos(\nu_j t)P_j^c + \sin(\nu_j t)R_j^c = T_j(t)$ *is similar to an orthogonal transformation on* V_j^c (*by a similarity transformation independent of* t) *and satisfies* $T_j(s + t) = T_j(s)T_j(t)$.

5. The Jordan and real canonical forms. The Jordan canonical form is derived from Theorem 1.3 by constructing a basis for each subspace V_i with respect to which the nilpotent transformation obtained by restricting the action of N_i to V_i has an appropriate representation. The standard construction of that basis is given in the next proposition.

5.1. PROPOSITION. *Let* V *be an* n*-dimensional* (*real or complex*) *vector space*, *and* $N\colon V \to V$ *be linear and nilpotent. Suppose that* $N^{k-1} \neq 0$ *but* $N^k = 0$. *For* $l = 1$ *to* k *let* $m_l = \dim\ker(N^l)$, *where* $\ker(N^l) = \{x \in V\colon N^l x = 0\}$ (*so that* $m_k = n$) *and let* $m_0 = 0$. *Then* $m_l - m_{l-1} \geqslant m_{l+1} - m_l$, *so that one can define integers* $b_l \geqslant 0$ *by*

$$b_k = m_k - m_{k-1}, \quad b_l = (m_l - m_{l-1}) - (m_{l+1} - m_l) \quad \textit{for } l = 1 \textit{ to } k - 1.$$

It is then possible to find, *for each* $l = 1$ *to* k, b_l *vectors* $\{v^{l,i}\}_{i=1}^{b_l}$ *so that* $B = \bigcup_{l=1}^{k}\bigcup_{i=1}^{b_l} Z_l(v^{l,i})$ (*where* $Z_l(v) = \{v, Nv, \ldots, N^{l-1}v\}$, *and* $N^l v^{l,i} = 0$) *is a basis for* V. *The matrix of* N *with respect to that basis has*, *for* $l = 1$ *to* k, *respectively*, b_l $l \times l$ *blocks along the diagonal each having the form* $[J_{pq}]$ *with* $J_{pq} = \delta_{p(q+1)}$, *and has zeroes elsewhere.*

Proof. It is evident that

$$\ker(N) \subset \ker(N^2) \subset \cdots \subset \ker(N^l) \subset \ker(N^{l+1}) \subset \cdots \subset \ker(N^k) = V.$$

We construct, in a particular way to be outlined below, a basis B_l for a subspace W_l of $\ker(N^l)$ such that $\ker(N^l) = W_l \oplus \ker(N^{l-1})$ $(l = 2$ to $k)$ and $\ker(N) = W_1$. Then automatically $B = \bigcup_{l=1}^{k} B_l$ will be a basis for $V = W_1 \oplus W_2 \oplus \cdots \oplus W_k$.

The construction of B_l (and the proof of the fact that $m_l - m_{l-1} \geqslant m_{l+1} - m_l$) proceeds by a reverse induction. We start by choosing a basis $B_k = \{v^{k,i}\}_{i=1}^{b_k}$ for a subspace W_k of $V = \ker(N^k)$ complementary to $\ker(N^{k-1})$. Suppose now that B_{l+1}, a basis for W_{l+1} (a subspace of $\ker(N^{l+1})$ complementary to $\ker(N^l)$) has been constructed. We prove that $N(B_{l+1})$ is an independent set in $\ker(N^l)$ and, in fact, that no non-trivial linear combination of vectors in $N(B_{l+1})$ can belong to $\ker(N^l)$. Suppose that $\Sigma c_i Nv^i \in \ker(N^{l-1})$, where $v^i \in B_{l+1}$. Then $\Sigma c_i v^i \in \ker(N^l)$, in which case all c_i's must be zero (since $\Sigma c_i v^i \in W_{l+1} \cap \ker(N^l) = \{0\}$ and the v^i's are independent). This implies that indeed

$$m_{l+1} - m_l = \dim W_{l+1} \leqslant m_l - m_{l-1}.$$

Moreover one can extend $N(B_{l+1})$ by b_l vectors to get a basis

$$B_l = N(B_{l+1}) \cup \{v^{l,i}\}_{i=1}^{b_l}$$

for a subspace W_l complementary to $\ker(N^{l-1})$ in $\ker(N^l)$. B_1 is defined by extending $N(B_2)$ to a basis of $\ker(N)$. That

$$B = \bigcup_{l=1}^{k} B_l = \bigcup_{l=1}^{k} \bigcup_{i=1}^{b_l} Z_l(v^{l,i})$$

is obvious on reflection.

Noting that

$$N(v^{l,i}) = Nv^{l,i}, N(Nv^{l,i}) = N^2 v^{l,i}, \ldots, N(N^{l-1}v^{l,i}) = 0$$

it is clear that each segment $Z_l(v^{l,i})$ of the basis corresponds to an $l \times l$ block of the specified form in the matrix representation of N.

Returning now to $A \in M_n(C)$ (or $A \in M_n(R)$ with all of its eigenvalues real) and to Theorem 1.3 (or Corollary 1.4), one uses the above Proposition to construct bases in each V_i corresponding to N_i. Letting S have the successive basis vectors as columns one has proved the following:

5.2. THEOREM. *Let $A \in M_n(C)$ (or $A \in M_n(R)$ with all its eigenvalues real). One can find $S \in M_n(C)$ (or $M_n(R)$) invertible, and such that $S^{-1}AS = J$, where J has Jordan blocks of the form $[\lambda_i \delta_{pq} + \delta_{p(q+1)}]$ along the diagonal, with zeroes elsewhere; the number of blocks of different sizes is determined by the numbers $m_{ij} = \dim\ker(N_i^j)$ and $n_i = \dim(V_i)$ as described in the previous proposition.*

One can similarly derive the real canonical form within the framework of Theorem 4.1.

5.3. THEOREM. *Let $A \in M_n(R)$. One can find $S \in M_n(R)$ invertible, such that $S^{-1}AS = C$, where along the diagonal C has Jordan blocks corresponding to the real variables λ_i $(i = 2m+1$ to $r)$ and real canonical blocks of the form*

$$\begin{bmatrix} A_j & O_2 & \cdot & \cdot & \cdot & \cdot & O_2 \\ I_2 & A_j & O_2 & \cdot & \cdot & \cdot & \cdot \\ O_2 & I_2 & A_j & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & O_2 & I_2 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & I_2 & A_j & O_2 \\ O_2 & \cdot & \cdot & \cdot & O_2 & I_2 & A_j \end{bmatrix}, \quad \text{with } I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad A_j = \begin{bmatrix} \mu_j & \nu_j \\ -\nu_j & \mu_j \end{bmatrix}, \quad O_2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

corresponding to complex pairs of eigenvalues $\mu_j \pm i\nu_j$.

Proof. We first choose a suitable basis for V_j^c, the range of $P_j^c = P_{2j-1} + P_{2j}$ in R^n. To do this note that once a suitable basis B_{2j-1} has been chosen for V_{2j-1} (the range of P_{2j-1} in C^n) using N_{2j-1} and Proposition 4.1, $B_{2j} = \overline{B}_{2j-1}$ is a basis for V_{2j}. Now it is an easy exercise to

check that the real and imaginary parts of the vectors in B_{2j-1} form a basis B_j^c for V_j^c. To find the matrix representation of A restricted to V_j^c we have to look at this more carefully.

Suppose $Z_l^{2j-1}(x+iy) = \{N_{2j-1}^p(x+iy)\}_{p=0}^{l-1}$ forms part of the basis B_{2j-1}. Set $x_p + iy_p = N_{2j-1}^{p-1}(x+iy)$. Then $\{x_1, y_1, x_2, y_2, \dots, x_l, y_l\}$ forms a segment of the basis of B_j^c. Notice that since $x_p + iy_p \in V_{2j-1}$ one has

$$A(x_p + iy_p) = (\mu_j + i\nu_j)(x_p + iy_p) + N_{2j-1}(x_p + iy_p) \quad \text{for } 1 \leqslant p < l$$

and

$$A(x_l + iy_l) = (\mu_j + i\nu_j)(x_l + iy_l).$$

Then

$$Ax_p = \mu_j x_p - \nu_j y_p + x_{p+1}, \quad Ay_p = \nu_j x_p + \mu_j y_p + y_{p+1} \quad \text{for } p = 1 \text{ to } l-1$$

and

$$Ax_l = \mu_j x_l - \nu_j y_l, \quad Ay_l = \nu_j x_l + \mu_j y_l.$$

The part of the matrix of A corresponding to this segment of the basis B_j^c, has the form of a real canonical block.

To define S one first uses the above procedures to obtain a basis for each B_j^c arranged in segments as above, and then finds bases B_i for V_i as in Proposition 4.1 (these are immediately real) and then defines S to have as columns the successive basis vectors in

$$B = \bigcup_{j=1}^{m} B_j^c \bigcup_{i=2m+1}^{r} B_i.$$

References

1. Richard Bronson, Differential Equations, Schaum's Outline Series, McGraw-Hill, 1968.
2. Morris W. Hirsch and Stephen Smale, Differential Equations, Dynamical Systems and Linear Algebra, Academic Press, 1974.
3. Seymour Lipschutz, Linear Algebra, Schaum's Outline Series, McGraw-Hill, 1968.

163.

MISCELLANEA

A Prediction

The solũtions (x, y, z) that satisfy an equation:

$$(1^n + 2^n + \cdots + x^n) + (1^n + 2^n + \cdots + y^n) = 1^n + 2^n + \cdots + z^n$$

$(x, y, z \in N;\ n \geqslant 2, n \in N)$ don't exist.

Nobuhisa Abe
I-I2
Nakakasuga-Town
Oita-City
Oita-Prefecture
Japan

Three people, but only two names, associated with a famous analytic collaboration. (See p. 214.)

Briefly, then, we must consider a general question: Is a mathematics teacher entitled to give inadequate arguments? I believe that there is indeed a place for them, though it is a strictly limited place. I do *not* mean to suggest that anyone should give a *fallacious* argument; that is never justified. But there is a difference between a false argument and an incomplete argument. An incomplete argument, I think, is justifiable under the following conditions:

(1) the omitted material is incidental, not central to the topic under discussion;
(2) the audience through its previous training is inclined to believe the statement whose proof is omitted; and
(3) the instructor really does know how to supply what has been left out.
I believe the following discussion qualifies under these standards.

There is no special role for 2 in this argument; what we want to show is that the square root of an integer n is irrational unless it is actually equal to another integer. The first pedagogical problem (unfamiliarity with proof by contradiction) can be met by postponing the "contradiction" to the very end, so that the students do not have to keep it in mind while also trying to understand other arguments. Thus it is easier if we phrase the main idea in "positive" language as follows:

If b/c is an "honest" fraction (not a whole number), then $(b/c)^2$ is also an honest fraction.

This is the sort of statement that any student can understand, since it can be checked in specific cases. It does of course imply the irrationality, since it shows that $\sqrt{n}$ cannot be an honest fraction; but that comment can be left to the end.

The proof of this main idea can now be described in a way that requires nothing but the sort of fraction manipulation with which all students will be familiar. They will understand the idea that we can reduce b/c to lowest terms, so that b and c have no factors in common. (The wording here is deliberately left vague.) But now when we form

$$(b/c)^2 = b^2/c^2 = (b \cdot b)/(c \cdot c),$$

the only factors that we have in the numerator are repetitions of the ones we had in b. Similarly, the only factors that we have in the denominator are repetitions of the ones we had in c. Hence again the numerator and denominator have no factors in common that could be cancelled, and we have an honest fraction. Thus the proof is finished.

Obviously the principles stated earlier imply that one should not give an argument like this in a course on theoretical arithmetic; the unique factorization properties that are slurred over here are a central topic in such a course and would have to be made explicit. Indeed, a lecturer on number theory might well choose to prove more generally, as in [**1**, p. 41] or [**2**, p. 112], that no monic equation with integer coefficients can have an honest fraction as a root. But when the irrationality of $\sqrt{2}$ is simply an illuminating sidelight on the main topic, an approach like the one I have described seems to be an effective compromise between unsupported assertion and belabored proof.

References

1. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, Oxford University Press, Oxford, 1960.
2. I. Niven and H. Zuckerman, An Introduction to the Theory of Numbers, Wiley, New York, 1972.
3. H. M. Stark, An Introduction to Number Theory, Markham, Chicago, 1970.

ANSWER TO PHOTO ON PAGE 185

J. D. Tamarkin, K. Hille, and E. Hille.

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy. Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

SOME SUSPICIOUSLY SIMPLE SEQUENCES

RICHARD K. GUY

Three problems of varying ages that have been drawn to our attention lately.

Donald E. G. Malm, Department of Mathematical Sciences, Oakland University, Rochester, MI 48063 asks about

Hofstadter's meta-Fibonacci sequence

Doug Hofstadter [**5**; see also **1**, **2**] has proposed the sequence defined by

$$Q(1) = Q(2) = 1, \quad Q(n) = Q(n - Q(n-1)) + Q(n - Q(n-2)), \quad n > 2.$$

If you calculate a few values:

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$Q(n)$	1	1	2	3	3	4	5	5	6	6	6	8	8	8	10	9	10	11	11	12
$Q(n+20)$	12	12	12	16	14	14	16	16	16	16	20	17	17	20	21	19	20	22	21	22
$Q(n+40)$	23	23	24	24	24	24	24	32	24	25	30	28	26	30	30	28	32	30	32	32
$Q(n+60)$	32	32	40	33	31	38	35	33	39	40	37	38	40	39	40	39	42	40	41	43
$Q(n+80)$	44	43	43	46	44	45	47	47	46	48	48	48	48	48	48	64	41	52	54	56

you are first struck by the monotonicity, though this breaks down from $Q(15)$ to $Q(16)$. Then from $Q(24)$ to $Q(25)$ it drops by 2; from $Q(31)$ to $Q(32)$ by 3; from $Q(48)$ to $Q(49)$ by 8; from $Q(96)$ to $Q(97)$ by 23 and from $Q(192)$ to $Q(193)$ by 56.

Do these drops become arbitrarily large?

In the main, the values of $Q(n)$ are a little larger than $\frac{1}{2}n$. New maxima include

$$Q(3 \times 2^k) = 2^{k+1}, \quad 0 \leqslant k \leqslant 8.$$

Does this continue? No! $Q(3 \times 2^9) = 808$, a far cry from 2^{10}.

You may also notice strings of constant values,

$$Q(3 \times 2^k - j) = 3 \times 2^{k-1} \quad (1 \leqslant j \leqslant k+1).$$

This is true for $1 \leqslant k \leqslant 5$, but for $k = 6$, $j = 7$, we find that $Q(185) = 94$, not 96, and the pattern soon disappears without trace. The sequence $Q(2^k + 1)$ is intriguing. In the following table:

k	0	1	2	3	4	5	6	7	8
$2^k + 1$	2	3	5	9	17	33	65	129	257
$Q(2^k + 1)$	1	2	3	6	10	17	31	57	106
e_k		0	1	0	2	3	3	5	8

the last line is calculated by $e_k = 2Q(2^{k-1} + 1) - Q(2^k + 1)$. The appearance of Fibonacci numbers is remarkable, though the reversal of 0 and 1 and the repetition of 3 do not inspire confidence. However, a similar table for $Q(3 \cdot 2^k + 1)$:

k	0	1	2	3	4	5	6	7	8
$3 \cdot 2^k + 1$	4	7	13	25	49	97	193	385	769
$Q(3 \cdot 2^k + 1)$	3	5	8	14	24	41	72	129	235
f_k		1	2	2	4	7	10	15	23
$\Delta f_k = e_k$?			1	0	2	3	3	5	8

in which $f_k = 2Q(3 \cdot 2^{k-1} + 1) - Q(3 \cdot 2^k + 1)$, yields, for the first differences of f_k, exactly the same sequence, e_k!! But if you continue either table, the pattern (such as it is) is rudely shattered in each case, and in quite different ways.

Hofstadter's original question was:

Is $Q(n)$ defined for all positive integers n?
I.e., does $Q(n)/n$ never exceed 1?

Malm has calculated the first 200,000 members of the sequence. He found only 27 + 11 values of n for which $Q(n)/n$ lies outside the interval [0.4, 0.6]. $Q(n)/n > 0.6$ for

$$
\begin{array}{rcccccccccccccc}
n = 1 & 3 & 4 & 6 & 7 & 8 & 9 & 12 & 13 & 15 & 18 & 24 & 31 & 48 & 63 \\
Q(n) = 1 & 2 & 3 & 4 & 5 & 5 & 6 & 8 & 8 & 10 & 11 & 16 & 20 & 32 & 40
\end{array}
$$

$$
\begin{array}{rccccccccccc}
n = 96 & 124 & 192 & 202 & 384 & 394 & 768 & 793 & 860 & 997 & 1545 & 1569 \\
Q(n) = 64 & 78 & 128 & 124 & 256 & 242 & 512 & 492 & 522 & 605 & 928 & 963
\end{array}
$$

and $Q(n)/n < 0.4$ for

$$
\begin{array}{rcccccccccc}
n = 193 & 385 & 395 & 433 & 769 & 794 & 801 & 846 & 1578 & 1673 & 1676 \\
Q(n) = 72 & 135 & 149 & 164 & 278 & 284 & 310 & 335 & 626 & 647 & 662
\end{array}
$$

It is remarkable that $Q(n)/n$ does not stray outside that interval between $n = 1676$ and $n = 200{,}000$. Moreover, in the range $10^5 < n \leqslant 2 \times 10^5$, Malm found only 77 values of n for which $Q(n)/n$ lies outside [0.45, 0.55], and they all occur in the range $100000 < n \leqslant 137958$. Malm later calculated the first million members and found more values of n for which $Q(n)/n$ is outside [0.45, 0.55]: 71 in the interval [202122, 278513], 44 in [394774, 556495] and 27 in [826568, 1000000].

Is there a sense in which $Q(n)$ is asymptotic to $\frac{1}{2}n$?
Do lim sup and lim inf of $Q(n)/n$ exist?

John Isbell, Department of Mathematics, SUNY at Buffalo, Buffalo, NY, 14214 has invented a game [4] and wants to know

Can Jack always beat the Giant at Beanstalk?

Beanstalk starts with the choice, by the Giant, of an odd integer, $n_0 > 1$. Thereafter Jack and the Giant play alternately according to the rule

$$
n_{i+1} = \begin{cases} n_i/2 & \text{if } n_i \text{ is even,} \\ 3n_i \pm 1 & \text{if } n_i \text{ is odd.} \end{cases}
$$

If n_i is even, there is only one option. If n_i is odd, there are just two. The winner is the player who arrives at 1.

This looks suspiciously like the notorious Collatz problem [3], [6]; does the fact that there is a choice make it any easier to analyse?

Suppose the Giant chooses 3. The play should go as follows, with Jack's moves in parentheses:

$$3 \quad (10) \quad 5 \quad (16) \quad 8 \quad (4) \quad 2 \quad (1)$$

and Jack wins. The only alternatives are 3 (8) 4 (2) 1 and the Giant wins, or 3 (10) 5 (14) 7 leading to a much more complicated situation that we'll examine later.

It's clear that Jack wins if he plays to a power of 4, and loses if he plays to an odd power of 2. Jack need never lose! One of his options is always singly even, of shape $4k + 2$ ($k > 0$) after which the Giant has to play the odd number $2k + 1$ and Jack *keeps control*. However, if he does this, each pair of moves increases n_t to $n_{t+2} \approx 3n_t/2$ and the game will go on forever. His alternative will be a move to $2^{2e}d$ or $2^{2e+1}d$, where d is odd and $e > 0$. If $d = 1$, we know that the former is a good move, and the latter is a bad one. If $d > 1$, the former loses control and the latter retains it. So the only real choice is between $2d$ (**hiding strategy**) and $2^{2e+1}d$ ($e > 0$, **greedy strategy**) on those occasions that it occurs. The hiding strategy can't win, but neither is the greedy strategy always successful. In the following game, starting with 7, Jack has four places, underlined, where he can adopt the greedy strategy: the options at other places lose control.

7 (22) 11 (34) 17 (50) 25 (74) 37 (110) 55 (166) 83 ($\underline{248}$) 124 (62) 31 (94) 47 (142) 71 (214) 107 (322) 161 (482) 241 (722) 361 (1082) 541 ($\underline{1624}$) 812 (406) 203 ($\underline{608}$) 304 (152) 76 (38) 19 ($\underline{56}$) 28 (14) 7 ...

and we have completed a cycle of 44 moves. We leave the reader to confirm that Jack can win in 65 moves, by a judicious mixture of the hiding and greedy strategies.

But is this true for all odd $n_0 > 1$?

And if it is, how long does Jack take to win? It is true for $n_0 < 1000$. The longest game (assuming that Jack always adopts the best strategy) that we have found has 263 moves and starts with 747. However, since a complete analysis involves a tree search, it is easy to make mistakes.

Isbell, Selfridge, Erdős and others have reminded me from time to time of an old problem considered by Mahler and Popken [7]:

What is the least number of ones needed to represent n, using only + and × (and parentheses)?

If we denote this number by $f(n)$, then, since the last operation is either multiplication or addition, we can define $f(n)$ recursively by the formula

$$f(n) = \min_{\substack{d \mid n \\ 2 \leqslant d \leqslant \sqrt{n} \\ 1 \leqslant e \leqslant n/2}} \{ f(d) + f(n/d), f(e) + f(n-e) \}.$$

However, this is somewhat less than useful. When the minimum is realized by the second member, is e ever greater than 1? Isbell and Myerson suggest that $n = (3^{70} + 3^{30} + 3^{10} + 1)^2 + 6$ might be such an example, with ¿$f(n) = 431$?

Is it always true that, for a prime p, $f(p) = 1 + f(p-1)$? And $f(2p) = \min\{2 + f(p), 1 + f(2p-1)\}$? The minimum is usually attained by the former, but also occasionally by the latter, e.g., for $p = 23$, 41 and 59. Is it always true that

$$3 + f(p) \leqslant 1 + f(3p - 1)?$$

Here is a table of the first 100 values of $f(n)$:

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$f(n)$	1	2	3	4	5	5	6	6	6	7	8	7	8	8	8	8	9	8	9	9
$f(n+20)$	9	10	11	9	10	10	9	10	11	10	11	10	11	11	11	10	11	11	11	11
$f(n+40)$	12	11	12	12	11	12	13	11	12	12	12	12	13	11	12	12	12	13	14	12
$f(n+60)$	13	13	12	12	13	13	14	13	14	13	14	12	13	13	13	13	14	13	14	13
$f(n+80)$	12	13	14	13	14	14	14	14	15	13	14	14	14	15	14	13	14	14	14	14

Selfridge uses induction to show that 3^k is the largest solution of $f(x) = 3k$, and that $3^k \pm 3^{k-1}$ are the largest solutions of $f(x) = 3k \pm 1$. If we write lb and lt for **binary** and **ternary log**, i.e., logs to base 2 and 3 respectively, we have

$$f(n) \geqslant 3 \operatorname{lt} n.$$

In fact $f(2^a 3^b) = 2a + 3b$ for $a = 0, 1$ and 2.

For what larger a is this true?

It is true for $2^a 3^b \leqslant 216$. But $2^a 3^b$ is not the largest n for which $f(n) = 2a + 3b$. For example $f(96) = 13$, but so is $f(108)$.

Are larger n always of the form $2^{a-3c}3^{b+2c}$?

Selfridge asks:

Is there an a for which $f(2^a) < 2a$?

We could find such a 2^a if its $\lceil a \operatorname{lt} 2 \rceil$ ternary digits had an average size of less than $2/\operatorname{lt} 2 - 3 \approx 0.1699250014$, but the expectation of this happening must be very small.

If we write n in binary, $n = \sum_{i=0}^{k} a_i 2^i$, where $0 \leqslant a_i \leqslant 1 = a_k$, then

$$n = a_0 + (1+1)(a_1 + (1+1)(a_2 + (\cdots (1+1)(a_{k-1} + 1 + 1) \cdots))),$$

where there are k parentheses $(1 + 1)$ and d ones, where $d(\leqslant k)$ is one less than the number of digits 1 in the binary expansion, $d = a_0 + a_1 + \cdots + a_{k-1}$, and

$$f(n) \leqslant 2k + d \leqslant 3k \leqslant 3 \operatorname{lb} n < 4.755 \operatorname{lt} n,$$

since $n \geqslant 2^k$.

Isbell notes that, for a set of density 1, namely those n which are nearly normal in the scale of 6,

$$f(n) < \left(\frac{19}{3} + \varepsilon\right) \log_6 n < 3.8833 \operatorname{lt} n.$$

Here $19/3$ is the average number of ones needed to convert an expression for k into an expression for $6k + l$, $0 \leqslant l \leqslant 5$. Similarly, working in duodecimal, there's a set of density 1 for which

$$f(n) < \left(\frac{26}{3} + \varepsilon\right) \log_{12} n < 3.8317 \operatorname{lt} n.$$

Some other bases $2^a 3^b$ will yield further small improvements, e.g., base 24 gives $f(n) < 3.817 \operatorname{lt} n$ for almost all n.

There are conflicting conjectures:

¿ For large n, $(3 + \varepsilon) \operatorname{lt} n$ ones suffice ?

¿ There are infinitely many n, perhaps a set of positive density for which $(3 + c) \operatorname{lt} n$ ones are needed, for some $c > 0$?

References

1. P. Erdős and R. L. Graham, Old and New Problems and Results in Combinatorial Number Theory, L'Enseignement Math., Geneva, 1980, p. 83.

2. Richard K. Guy, Unsolved Problems in Number Theory, Springer-Verlag, New York, 1981, Problems E31 and F26.

3. Richard K. Guy, Don't try to solve these problems! this MONTHLY, 90 (1983) 35–41.

4. Richard K. Guy, John Isbell's game of Beanstalk and John Conway's game of Beans-Don't-Talk, Math. Mag.

5. Douglas Hofstadter, Gödel, Escher, Bach: an Eternal Golden Braid, Basic Books, New York, 1979, p. 137.

6. Jeffery C. Lagarias, The $3x + 1$ problem and its generalizations, this MONTHLY, 92 (1985) 3–23.

7. K. Mahler and J. Popken, On a maximum problem in arithmetic (Dutch), Nieuw Arch. Wiskunde, (3) 1 (1953) 1–15; MR 14, 852e.

NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.

For instructions about submitting Notes for publication in this department see the inside front cover.

THE TIETZE EXTENSION THEOREM AND THE OPEN MAPPING THEOREM

SANDY GRABINER
Department of Mathematics, Pomona College, Claremont, CA 91711

The Open Mapping Theorem says that a bounded linear transformation from one Banach space onto another must be an open mapping, while the Tietze Extension Theorem says that a bounded continuous function can always be extended from a closed subset of a normal topological space to the entire space. The two theorems sound as though they are unrelated, but the central arguments in the standard proofs of these two theorems are in fact quite similar. In each proof we can first crudely approximate a continuous function or an element of an abstract Banach space. We then inductively construct better approximations, by, at each stage, approximating the error from the previous stage. This process produces an infinite series that converges exactly to the original element. For the Open Mapping Theorem, one proves that if T is a bounded linear operator and B is an open ball about the origin for which the closure $\overline{T(B)}$ contains a ball about the origin, then $T(B)$ already contains such a ball. In fact, as many people have noticed, the usual proof does not use the fact that $T(B)$ is actually dense in some ball, but only that it gets "close enough" to elements in this ball. The following lemma gives a precise statement of what one can prove. Furthermore, this precise statement is strong enough to give the Tietze Extension Theorem.

APPROXIMATION LEMMA. *Suppose that T is a bounded linear operator from a (real or complex) Banach space E to a Banach space F, and suppose also that m and r are positive numbers with $r < 1$. If for each y in F, there is an x_0 in E with $\|x_0\| \leqslant m\|y\|$ and $\|y - Tx_0\| \leqslant r\|y\|$, then there is also an x in E with $Tx = y$ and $\|x\| \leqslant m\|y\|/(1 - r)$.*

Proof. For simplicity take $\|y\| = 1$. Applying the hypothesis to $y - Tx_0$ in place of y, we find an x_1 with $\|x_1\| \leqslant rm$ and $\|y - T(x_0 + x_1)\| \leqslant r^2$. Proceeding inductively, we obtain a sequence $\{x_n\}_0^\infty$ with

$$\|x_n\| \leqslant r^n m, \tag{1}$$

and

$$\|y - T(x_0 + x_1 + \cdots x_n)\| \leqslant r^{n+1}. \tag{2}$$

Formula (1) implies that the series $\sum_0^\infty x_n$ converges absolutely to a vector x of norm less than or equal to $m/(1 - r)$. Letting n go to infinity in formula (2) then shows that $y = Tx$.

TIETZE EXTENSION THEOREM. *If M is a closed subset of a normal space X, then any bounded continuous real-valued function on M may be extended to a continuous function on X with the same bound.*

Proof. Let T be the restriction map from the space of bounded continuous functions on X to the space of bounded continuous functions on M. When these function spaces are given the sup norms, T becomes a bounded operator. We will verify that T satisfies the hypothesis of the Approximation Lemma with $m = 1/3$ and $r = 2/3$ by repeating the first step of the standard proof of the Tietze Extension Theorem [**1**, Theorem 3.2, p. 212]. Suppose that g is a continuous function on M with sup norm 1. Let $A = g^{-1}[-1, -1/3]$ and $B = g^{-1}[1/3, 1]$. By Urysohn's lemma, there is a continuous function f from X to $[-1/3, 1/3]$ with f identically equal to $-1/3$ on A and to $1/3$ on B. Then $\|f\| = 1/3$ and $\|Tf - g\| \leqslant 2/3$. This completes the proof.

Another form of the Approximation Lemma [**2**, Lemma 4.13(a), p. 94–95], which is useful in operator theory, states that if B_1 and B_2 are open balls about the origins in E and F, respectively, and if $\overline{T(B_1)} \supseteq B_2$, then $T(B_1) \supseteq B_2$. This follows easily by applying the form of the Approximation Lemma proved above to a sequence $\{r_n\}$ with limit 0.

References

1. J. R. Munkres, Topology: A First Course, Prentice-Hall, Englewood Cliffs, NJ, 1975.
2. Walter Rudin, Functional Analysis, McGraw-Hill, New York, 1973.

THE PROBABILITY OF HEADS*

JOSEPH B. KELLER
Departments of Mathematics and Mechanical Engineering, Stanford University, Stanford, CA 94305

1. Introduction. Why is the outcome of a coin toss considered to be random, even though it is uniquely determined by the laws of physics and the initial conditions? If it is random, why is there a definite probability associated with each outcome, regardless of how the coin is tossed? Finally, if there is a definite probability for each outcome, how can it be calculated? We shall try to answer these questions by analyzing the motion of a tossed coin. Then we shall extend our considerations to a wheel and other "chance" devices.

2. Mechanics of a tossed coin. Let us consider a circular coin of radius a and negligible thickness, one side of which is marked heads and the other side tails. We assume that the center of gravity of the coin is at its geometrical center, the height of which we denote $y(t)$ at time t. Then Newton's equation for the vertical motion of the center of gravity of the coin is

$$\frac{d^2y(t)}{dt^2} = -g. \tag{1}$$

Here the positive constant g is the acceleration of gravity. To supplement (1) we suppose that initially, at time $t = 0$, the center of the coin is at height a and that it has an upward velocity u. Thus we have the initial conditions

$$y(0) = a, \quad \frac{dy(0)}{dt} = u. \tag{2}$$

The differential equation (1) and the initial conditions (2) determine $y(t)$. Instead of $y(0) = a$ we could have prescribed any other initial value. This particular choice will simplify some of the subsequent calculations.

*Research supported by the Office of Naval Research, the Air Force Office of Scientific Research, the Army Research Office, and the National Science Foundation.

In addition to its vertical motion, the coin is assumed to be rotating about a horizontal axis that lies along a diameter of the coin. We shall choose the z-axis to be parallel to this rotation axis. Then we can describe the angular position of the coin at time t by the angle $\theta(t)$ between the positive y-axis and the normal to the side of the coin marked heads, both of which lie in the x, y plane. (See Fig. 1.) Now the equation governing the rotational motion of the coin is simply

$$\frac{d^2\theta(t)}{dt^2} = 0. \tag{3}$$

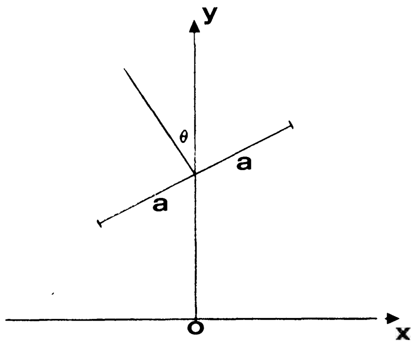


FIG. 1. The x, y plane intersects the coin along a diameter of length $2a$. The normal to the side of the coin marked heads makes the angle θ with the positive y-axis.

We assume that initially the coin is horizontal with side heads up, so that $\theta(0) = 0$, and that it has the positive angular velocity ω. Thus we specify the initial conditions

$$\theta(0) = 0, \quad \frac{d\theta(0)}{dt} = \omega. \tag{4}$$

Here also we could have replaced $\theta(0) = 0$ by any other initial value, but this choice will simplify our calculations.

The solutions of (1) and (2) for $y(t)$ and of (3) and (4) for $\theta(t)$ are

$$y(t) = ut - \frac{gt^2}{2} + a, \tag{5}$$

$$\theta(t) = \omega t. \tag{6}$$

These equations hold from $t = 0$ until the first time $t_0 > 0$ at which the coin lands on a surface, which we take to be the plane $y = 0$. We shall assume that whichever side of the coin is up at t_0 remains up. This will be the case if the coin lands in sand or mud, but not if it lands on a hard surface where it will bounce, roll, etc. Thus the coin will end its motion with heads up if

$$2n\pi - \frac{\pi}{2} < \theta(t_0) < 2n\pi + \frac{\pi}{2}, \quad n = 0,1,2,\dots. \tag{7}$$

To find t_0, we consider the lowest point of the coin at time t, which is at $y(t) - a|\sin\theta(t)|$. Then t_0 is the smallest positive root of the equation

$$y(t_0) - a|\sin\theta(t_0)| = 0. \tag{8}$$

3. The pre-image of heads. We shall now analyze (7) and (8) to find out if the coin ends up heads for given values of the initial velocity u and the initial angular velocity ω. The set of all pairs u, ω of nonnegative values for which it ends up heads, we shall call the pre-image of heads in the u, ω plane, and we shall denote it H. First we consider the end points of the intervals in (7), which are given by $\theta(t_0) = (2n \pm \frac{1}{2})\pi$. At them (6) yields

$$\omega t_0 = \left(2n \pm \frac{1}{2}\right)\pi. \tag{9}$$

Now since $\sin\theta(t_0) = \pm 1$, (8) becomes $y(t_0) - a = 0$. When (5) is used in this equation it becomes

$$ut_0 - gt_0^2/2 = 0. \tag{10}$$

The positive solution of (10) is $t_0 = 2u/g$. Then we use this result in (9) to obtain

$$\omega = \left(2n \pm \frac{1}{2}\right)\frac{\pi g}{2u}, \quad n = 0,1,2\dots. \tag{11}$$

The relation (11) corresponds to the endpoints of the intervals in (7), and therefore it determines the boundaries of the region H in the u, ω plane. This relation is graphed in Fig. 2 for many values of n. Each curve is a hyperbola. On the axis $\omega = 0$ heads remains up throughout the toss, so this axis and the adjacent strip lie in H. The next strip lies in T, the pre-image of tails, and the successive strips alternate between H and T, as we can see by examining (7).

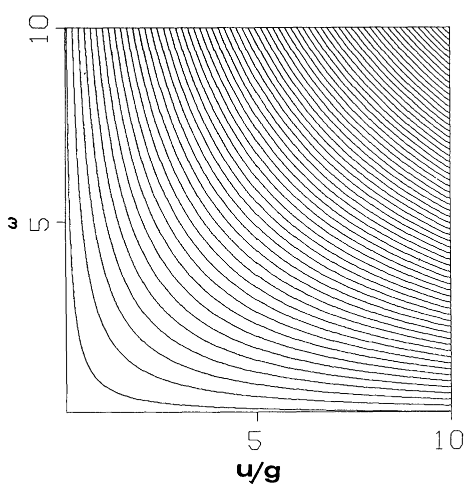


FIG. 2. The curves which separate the sets H and T, the preimages of heads and tails in the u, ω plane of initial conditions, are shown for various values of n. These are based upon (11), with the abscissa being u/g. The lowest strip, adjacent to the axes, belongs to H, the next to T, and so on alternately.

From (11) we find that the vertical separation between any two adjacent boundary curves is $\pi g/2u$, except that the lowest one is only $\pi g/4u$ above the axis. Thus the strips are of equal vertical width, and this width tends to zero as u increases.

In order to find where actual tosses lie on this figure, we can consider the maximum height h to which a coin rises. From (5) we find that $dy/dt = 0$ at $t_m = u/g$, and that

$$h = y(t_m) = u^2/2g + a.$$

Thus $u = [2g(h-a)]^{1/2}$. Now $g = 32$ feet per second^2, so if $h - a = 1$ foot we find $u = 8$ ft/sec and $t = u/g = 1/4$ sec.

To find ω Professor Persi Diaconis observed a number of typical tosses of a coin under stroboscopic illumination, and found that $\omega \approx 38$ revolutions/sec $= 38(2\pi)$ radians/sec. The number of revolutions per toss is

$$n = \omega t_0 = 2u\omega/g \approx \frac{2}{4}(38) \text{ revs/toss} = 19 \text{ revs/toss}.$$

Thus the point $(u/g, \omega) = ((1/4)$ sec, $76\ \pi/$sec$)$ is way above the region shown in Fig. 2. It lies near the lines corresponding to $n = 19$ in (9).

4. The probability of heads. So far we have determined the sets H and T in the u, ω plane, which are the pre-images of heads and tails respectively. Now we suppose that the initial condition

u, ω is a random variable with a continuous probability density $p(u, \omega)$ with support in the region $u > 0$, $\omega > 0$. Then the probability of heads P_H is given by

$$P_H = \iint_H p(u, \omega)\, du\, d\omega. \tag{12}$$

Thus the outcome is random because we have assumed that the initial conditions are random. This is the answer to the first question we asked in the introduction.

For any value P_H in the interval zero to one, there are densities p for which the integral in (12) has that value. Thus (12) does not seem to place any restriction on P_H. However, we shall now show that when the support of p is shifted to sufficiently large values of u, and possibly to large values of ω also, then P_H tends to a fixed limit which is independent of the density $p(u, \omega)$. This is the content of the following theorem.

THEOREM 1. *Let $p(u, \omega)$ be a continuous probability density with support in the region $u > 0$, $\omega > 0$, and let β be a fixed constant satisfying $0 \leqslant \beta \leqslant \pi/2$. Then*

$$\lim_{U \to \infty} P_H = \lim_{U \to \infty} \iint_H p\left(u - U\cos\beta, \omega - a^{-1}U\sin\beta\right) d\omega\, du = \frac{1}{2}. \tag{13}$$

Proof. The proof is given in the appendix, although the conclusion is evident from Fig. 2.

The significance of this theorem is that there is a unique probability of heads which is approximately achieved by any continuous probability density of the initial values u, ω that is shifted to sufficiently large values of u and ω. The approximation improves as the density $p(u, \omega)$ is translated to larger values of u and ω.

The limiting value of P_H is $1/2$, despite the fact that the initial condition is not symmetric in heads and tails, since the coin always starts out with heads up. Therefore the traditional method of calculating P_H, based upon symmetry, is not applicable.

The reason why P_H has a limit as U increases is that the pre-images H and T both consist of many strips which become very narrow at infinity. Thus both H and T occupy fixed fractions of the area of any disk which is shifted to infinity. This answers the second question in the introduction. By calculating those fractions we get the limiting values of P_H and P_T, which answers the third question in the introduction.

5. Wheels. Another common gambling device, which is often used in carnivals, is a rotating wheel with numbers marked along its outer edge, and a pointer. The wheel is spun and ultimately comes to rest, with the number indicated by the pointer being the winning number. Usually there are nails between adjacent numbers which are hit by the pointer, to aid in slowing down and stopping the wheel, and to make clear which number is indicated by the pointer when the wheel stops. We shall ignore the nails, and analyze the motion of the wheel.

The angular position of the wheel can be described by specifying the angular distance θ from the pointer to some mark on the wheel. Let $\theta(t)$ be this angle at time t. To determine $\theta(t)$, we assume that there is a constant frictional torque retarding the motion, and that this frictional torque vanishes when the wheel comes to rest. Then while the wheel is turning in the counterclockwise direction, so that $\theta(t)$ is increasing with t, its equation of motion is

$$\frac{d^2\theta(t)}{dt^2} = -\alpha, \quad 0 \leqslant t < t_0. \tag{14}$$

Here α is a positive constant equal to the retarding torque divided by the moment of inertia of the wheel, and t_0 is the first time at which the wheel comes to rest. Thus

$$\frac{d\theta(t_0)}{dt} = 0. \tag{15}$$

We assume that initially the mark is under the pointer, so that $\theta(0) = 0$, and that the wheel has an initial angular velocity $\omega > 0$. Thus

$$(16) \qquad \theta(0) = 0, \quad \frac{d\theta(0)}{dt} = \omega.$$

The solution of (14) and (16) is

$$(17) \qquad \theta(t) = \omega t - \alpha t^2/2, \quad 0 \leqslant t \leqslant t_0.$$

To find t_0 we use (17) in (15) to get $\omega - \alpha t_0 = 0$. Thus $t_0 = \omega/\alpha$ and (17) yields

$$(18) \qquad \theta(t_0) = \frac{\omega^2}{2\alpha}.$$

Now (18) gives the final position of the mark, since the wheel stops moving at t_0.

Let us now solve (18) for ω to find which points on the positive ω-axis correspond to a given final position θ. To do so we set $\theta(t_0) = \theta + 2n\pi$ in (18) with $n = 0, 1, 2 \ldots$, since all such coordinates represent the same position of the wheel. Then the solution of (18) yields

$$(19) \qquad \omega_n = [2\alpha(\theta + 2n\pi)]^{1/2}, \quad n = 0, 1, 2, \ldots .$$

This set is the pre-image of θ on the positive ω-axis, which is the set of initial conditions. From (19) we find that for large n the spacing between successive values of ω_n is

$$(20) \qquad \omega_{n+1} - \omega_n = \frac{2\alpha\pi}{\omega_n}[1 + O(\omega_n^{-2})].$$

Thus the ω_n become ever more closely spaced as n increases.

Now we suppose that ω is random with a continuous probability density $p(\omega)$ which vanishes for $\omega < 0$. Then the corresponding probability density $P(\theta)$ of θ is

$$(21) \qquad P(\theta) = \sum_{n=0}^{\infty} p[\omega_n(\theta)] \frac{d\omega_n(\theta)}{d\theta} = \alpha \sum_{n=0}^{\infty} p[\omega_n(\theta)] / \omega_n(\theta).$$

When $p(\omega)$ is shifted to larger values of ω by the amount Ω, (21) yields

$$(22) \qquad P_\Omega(\theta) = \alpha \sum_{n=0}^{\infty} p[\omega_n(\theta) - \Omega] / \omega_n(\theta).$$

In the appendix we obtain from (22) the following result:

THEOREM 2. *If $p(\omega)$ is a continuous probability density which vanishes for $\omega < 0$, then*

$$(23) \qquad \lim_{\Omega \to \infty} P_\Omega(\theta) = \frac{1}{2\pi}.$$

Thus in this case also there is a unique probability distribution of the outcome that results approximately from any initial density $p(\omega)$ which has its support shifted to sufficiently large values of ω. Again the limit distribution is what would be given by a symmetry argument, although the initial condition is not symmetric because $\theta(0) = 0$.

6. Other chance devices. We shall now consider any mechanical device used in a game of chance. We assume that its motion is determined by the laws of mechanics and its initial condition, which we denote u. We also assume that its possible final states can be partitioned into a finite collection of subsets $S_1, S_2, \ldots S_n$, each of which we identify with one outcome or event also denoted $S_1, S_2, \ldots S_n$. The laws of mechanics determine a unique final state corresponding to each initial condition u. All those initial conditions which lead to a final state in the set S_i we call the pre-image of the outcome S_i. We denote the pre-image of S_i by H_i.

Now we suppose that there is a probability density $p(u)$ defined on the space of initial conditions. Then the probability P_i of outcome S_i is given by

(24) $$P_i = \int_{H_i} p(u)\, du.$$

The question we consider is "When is P_i independent of the particular probability density $p(u)$; and if it is independent, what is the value of P_i?"

The preceding examples of coin tossing and wheel spinning suggest an answer. It is to consider the behavior of the H_i at infinity in the space of initial conditions. Suppose that a fixed fraction f_i of the volume of any small region at infinity is contained in H_i. Then it follows from (24) that P_i will have the limit f_i as $p(u)$ is shifted to infinity.

The meaning of this conclusion is that when the H_i have the required property, "any" probability density of initial conditions concentrated at very large initial conditions will yield nearly the same probabilities $P_i = f_i$ of the various outcomes. Whether or not the H_i do have the requisite property, and what the fractions f_i are, can be decided in principle by analyzing the mechanical behavior of the device.

7. Some related work. Poincaré [**1**] treated a model of the roulette wheel which led to results like those in Section 5, although it does not involve any mechanics. His results and some related ones are mentioned in Feller [**2**]. Smoluchowski [**3**] presented similar ideas. Then Hopf [**4**], [**5**] introduced concepts from mechanics into the calculation of probabilities, and a point of view just like the one we have described. Recently Thorp [**6**] has used mechanical considerations in analyzing roulette.

Acknowledgement. It is a pleasure to thank Professor Persi Diaconis for raising a question about the face probabilities of noncubical dice which led me to think about this subject, and for permitting me to quote his measurements of the rotation rates of tossed coins.

Appendix. We shall now indicate the proofs of the two theorems.

Proof of Theorem 1. We first write the integral in (13) as an iterated integral, and use (11) to determine the range of $\omega' = \omega - a^{-1}U\sin\beta$:

(A1) $$P_H = \int_{U\cos\beta}^{\infty} \sum_{n=0}^{\infty} \int_{(2n-1/2)\pi g/2u - a^{-1}U\sin\beta}^{(2n+1/2)\pi g/2u - a^{-1}U\sin\beta} p(u - U\cos\beta, \omega')\, d\omega'\, du.$$

The lower limit of u is $U\cos\beta$ because $p = 0$ for $u - U\cos\beta < 0$. As U tends to infinity, so does u provided that $\beta < \pi/2$. Then the range of each integral over ω' is of length $O(U^{-1})$. Therefore we can approximate each of these integrals by the length of the interval multiplied by the value of the integrand at the midpoint. The error in this approximation is $O(U^{-1})$. Thus we can rewrite (A1) as

(A2) $$P_H = \int_{U\cos\beta}^{\infty} \sum_{n=0}^{\infty} p\left(u - U\cos\beta, \frac{2n\pi g}{2u}\right)\frac{\pi g}{2u}[1 + o(1)]\, du.$$

As U becomes infinite, the sum in (A2) converges to one half the Riemann integral of p with respect to ω'. This is clear if p has compact support, and if not it follows by approximating p by a sequence of functions with compact support. Thus we have

(A3) $$P_H = \int_{U\cos\beta}^{\infty} \frac{1}{2}\int_0^{\infty} p(u - U\cos\beta, \omega')\, d\omega'\, du[1 + o(1)].$$

Upon setting $u' = u - U\cos\beta$ in (A3), and remembering that the integral of p is one, we get

(A4) $$P_H = \frac{1}{2}\int_0^{\infty}\int_0^{\infty} p(u', \omega')\, d\omega'\, du'[1 + o(1)] = \frac{1}{2}[1 + o(1)].$$

When U tends to infinity, $o(1)$ tends to zero, and (A4) yields the result in (13). For $\beta = \pi/2$ we introduce $u' = u - U\cos\beta$ instead of ω' and then the proof is similar to that above.

Proof of Theorem 2. By using (20) for ω_n^{-1} we can write (22) as

$$P_\Omega(\theta) = \frac{1}{2\pi} \sum_{n=0}^{\infty} p[\omega_n(\theta) - \Omega][\omega_{n+1}(\theta) - \omega_n(\theta)][1 + O(\omega_n^{-2})]. \tag{A5}$$

Then we write $w_n(\theta) = \omega_n(\theta) - \Omega$, and note that $p(\omega_n - \Omega) = 0$ for $\omega_n < \Omega$. Thus we can write (A5) as

$$P_\Omega(\theta) = \frac{1}{2\pi} \sum_{n=0}^{\infty} p[w_n(\theta)][w_{n+1}(\theta) - w_n(\theta)][1 + O(\Omega^{-2})]. \tag{A6}$$

Now $w_{n+1} - w_n = \omega_{n+1} - \omega_n = O(\omega_n^{-1})$, and $\omega_n > \Omega$ whenever $p(\omega_n - \Omega) > 0$. Therefore $w_{n+1} - w_n = O(\Omega^{-1})$, and the sum in (A6) converges to the Riemann integral of p as Ω becomes infinite. Since this integral is one, (A6) yields (23).

References

1. H. Poincaré, Calcul des Probabilités, Gauthier-Villars, Paris, 1912.
2. W. Feller, An Introduction to Probability Theory and its Applications, Vol. II, 2nd ed., Wiley, New York, 1971.
3. M. v. Smoluchowski, Naturwissenschaften, 6 (1918) 253.
4. E. Hopf, On causality, statistics and probability, J. Math. Phys., 13 (1934) 51–102.
5. ______, Remarks on causality and probability, J. Math. Phys., 14 (1935) 4–9.
6. E. O. Thorp, Physical prediction of roulette, (unpublished).

ANCESTORS, CARDINALS, AND REPRESENTATIVES

T. D. PARSONS
Department of Mathematics, California State University, Chico, CA 95929-0525

1. Introduction. Occasionally we come across a proof so charming that it is as memorable or more so than the theorem it proves. One such is the "parent-ancestor-descendant" proof of the Schroeder–Bernstein Theorem which appeared in Birkhoff and Mac Lane [**2**] and has ever since been a favorite of college mathematics audiences. This proof is a clever restatement of an argument of S. Banach [**1**], and it demonstrates the value of picturesque terminology in good mathematical exposition. In our expository note here, we apply the same proof to obtain a more general theorem (Theorem 1 below) and show how to derive as its corollaries not only Banach's mapping theorem and the Schroeder-Bernstein Theorem, but also some results of Mendelsohn and Dulmage [**6**], R. Rado [**7**, Theorem 4.3], and Hoffman and Kuhn [**5**] about systems of distinct representatives for families of sets.

Theorem 1 is due to Ore [**8**] and independently to Perfect and Pym [**9**], whose version we state first.

THEOREM 1. *Let $X' \subseteq X$, $Y' \subseteq Y$, and $E \subseteq X \times Y$ be sets. Suppose that there exist injective mappings $f: X' \to Y$ and $g: Y' \to X$ such that $(x, f(x)) \in E$ and $(g(y), y) \in E$ for all $x \in X'$ and $y \in Y'$. Then there exist sets X_0, Y_0 with $X' \subseteq X_0 \subseteq X$ and $Y' \subseteq Y_0 \subseteq Y$ and a bijection $h: X_0 \to Y_0$ such that $(x, h(x)) \in E$ for all $x \in X_0$.*

Ore's version, which is more convenient for certain applications, is stated (Theorem 1′) in the language of bipartite graphs. Here we consider X, Y to be disjoint sets of vertices forming the bipartition for the bipartite graph B with edge-set E—that is, vertex x is adjacent to vertex y if $(x, y) \in E$. A subset $S \subseteq E$ of edges is said to be *incident* to a subset X' of X (respectively, Y' of Y) if every element of X' (Y') is an end-point of some edge (x, y) of S. A set $M \subseteq E$ is *independent* if no two edges of M share a common endpoint (i.e., M is a "matching" in B).

THEOREM 1′. *Let $B = [X, Y]$ be a bipartite graph with edge-set $E \subseteq X \times Y$. Suppose that*

$X' \subseteq X$ and $Y' \subseteq Y$. If there is an independent set S incident to X' and an independent set T incident to Y', then there is an independent set M incident to each of X', Y'.

The only real difference between the statements of Theorem 1 and Theorem 1′ is in whether or not the sets X and Y must be taken to be disjoint. To obtain Theorem 1′ from Theorem 1, just define f and g by $(x, f(x)) \in S$ for all $x \in X'$ and $(g(y), y) \in T$ for all $y \in Y'$ and let $M = \{(x, h(x)) | x \in X_0\}$. Conversely, replacing Y by a copy disjoint from X, Theorem 1 follows easily from Theorem 1′.

2. Ancestors. We now prove Theorem 1.

We may assume without loss of generality that $X \cap Y = \varnothing$. If $g(y) = x$, then we call y the *parent* of x. Similarly, if $f(x) = y$, we call x the parent of y. An element z is descended from its parent (if any), its parent's parent (if any), and so forth, which are in turn its *ancestors*. Define sets P and Q by

$$P = \{ x \in X' | x \text{ is parentless or has a parentless ancestor in } X' \},$$

$$Q = \{ x \in g(Y') | x \text{ has no parentless ancestor, or } x \text{ has a parentless ancestor not in } X' \}.$$

Clearly, $P \cap Q = \varnothing$ and $P \cup Q \supseteq X'$. Let $X_0 = P \cup Q$ and define $h : X_0 \to Y$ by

$$h(x) = \begin{cases} f(x), & \text{if } x \in P, \\ g^{-1}(x), & \text{if } x \in Q. \end{cases}$$

Then h is a bijection from X_0 to $Y_0 = h(X_0) \subseteq Y$, and for every $x \in X_0$ we have $(x, h(x)) \in E$.

Let $y \in Y'$ and $x = g(y)$. If $x \in Q$, then $h(x) = g^{-1}(x) = y$. If $x \notin Q$, then x must have a parentless ancestor $x' \in X'$, which is also an ancestor of y (since $X \cap Y = \varnothing$ implies that $y \neq x'$). Therefore y must have a parent $x'' \in X'$ either equal to x' or descended from x', so that $x'' \in P$ and $h(x'') = f(x'') = y$. Therefore $Y' \subseteq h(X_0) = Y_0$, concluding the proof.

3. Cardinals. For any set S, we let $|S|$ denote the cardinality of S. Recall that $|S|$ stands for the equivalence class of S under the equivalence relation "$S \equiv T$ if and only if there is a bijection from S to T". We then compare cardinal numbers by letting $|X| \leqslant |Y|$ if there is an injective mapping from X to Y. The relation "$\leqslant$" here is obviously reflexive and transitive. The Schroeder–Bernstein Theorem says that it is also antisymmetric:

$$|X| \leqslant |Y| \quad \text{and } |Y| \leqslant |X| \quad \text{imply that } |X| = |Y|.$$

Banach's mapping theorem [1] states that if $f : X \to Y$ and $g : Y \to X$ are injective mappings, then there exist partitions $X = X_1 \cup X_2$ and $Y = Y_1 \cup Y_2$ such that $f(X_1) = Y_1$ and $g(Y_2) = X_2$. (This immediately implies the Schroeder–Bernstein Theorem.)

Theorem 1 reduces to Banach's mapping theorem when $X' = X$ and $Y' = Y$ [take $E = \{(x, f(x)) | x \in X\} \cup \{(g(y), y) | y \in Y\}$ and let $X_1 = \{x \in X | h(x) = f(x)\}$], and to the Schroeder–Bernstein Theorem when, further, $E = X \times Y$.

4. Representatives. Let I be an index set, and $F = \{Y_i | i \in I\}$ a family of subsets of a set Y. An injective mapping $y : I \to Y$ is a *system of distinct representatives* (SDR) for F if $y(i) \in Y_i$ for all $i \in I$. We may identify such a mapping with the indexed list $\{y(i) | i \in I\}$ of its images, which we also call an SDR for F.

If for every finite subset $K \subseteq I$ we have that $|\cup \{Y_i | i \in K\}| \geqslant |K|$, then we say that F satisfies the *condition H*. Clearly this condition is necessary for F to have an SDR. Under certain conditions it is also sufficient, such as when I is finite (P. Hall [4]) or when every Y_i is finite (Marshall Hall, Jr. [3]). In general condition H is not sufficient, as may be seen from examples such as $I = \{0, 1, 2, \dots\}$, $Y_0 = \{1, 2, 3, \dots\}$, $Y_i = \{i\}$ for $i = 1, 2, 3, \dots$, where no SDR exists.

Ore's Theorem 1′ may be used to "symmetrize" the setting for SDR's. A family $F = \{Y_x | x \in X\}$ of subsets of a set Y indexed by a set X, corresponds naturally to a bipartite graph $B = [X, Y]$

with edge-set

$$E = \{(x, y) | x \in X \text{ and } y \in Y_x\} \subseteq X \times Y.$$

If for each $y \in Y$ we let $X_y = \{x \in X | y \in Y_x\}$ be the neighbor-set of y in B, and we let $G = \{X_y | y \in Y\}$, then, symmetrically, G is a family of subsets of X indexed by Y and $E = \{(x, y) | y \in Y \text{ and } x \in X_y\}$.

Subsets $X' \subseteq X$ and $Y' \subseteq Y$ define subfamilies $F' = \{Y_x | x \in X'\}$ and $G' = \{X_y | y \in Y'\}$. Evidently, the indexed list $\{y(x) | x \in X'\}$ of elements in Y is an SDR for F' if and only if the set $S = \{(x, y(x)) | x \in X'\}$ is an independent set incident to X' relative to E. Similarly, an indexed list $\{x(y) | y \in Y'\}$ is an SDR of G' if and only if $T = \{(x(y), y) | y \in Y'\}$ is an independent set incident to Y'.

Thus, in the context of SDR's, Theorem 1$'$ says that if F' and G' each have SDR's then there is a bijection $h: X_0 \to Y_0$ (defined by $M = \{(x, h(x)) | x \in X_0\}$) for certain sets $X_0 \supseteq X'$ and $Y_0 \supseteq Y'$ which provides, simultaneously, the SDR's $\{h(x) | x \in X'\}$ and $\{h^{-1}(y) | y \in Y'\}$ for F' and G'.

When F' and G' are families of *finite* sets, then a necessary and sufficient condition for both families to have SDR's is that both families satisfy condition H (by the result of Marshall Hall [**3**]), so that such a bijection h providing simultaneous SDR's exists exactly in this case. Let us call this result *Corollary* 2, which is easier to restate graph theoretically as

COROLLARY 2$'$. *Let $B = [X, Y]$ be a bipartite graph. Suppose that $X' \subseteq X$ and $Y' \subseteq Y$ are subsets of vertices each having only finitely many neighbors in B. A necessary and sufficient condition that there exist a matching M incident to each of X' and Y' in B is that for every choice of finite subsets $K \subseteq X'$ and $L \subseteq Y'$, the vertices in K (respectively, L) have collectively at least $|K|$ ($|L|$) neighbors in Y (X).*

When X and Y are both finite sets, Corollary 2$'$ becomes a result of Mendelsohn and Dulmage [**6**, Theorem 1, p. 231], which in turn was a symmetrized form of a result of Hoffman and Kuhn [**5**, p. 456] corresponding to the further restriction $X' = X$. The Hoffman–Kuhn result was stated in terms of SDR's which contain a prescribed set of elements $Y' \subseteq Y$; such problems arise naturally from the set-theoretic interpretation Corollary 2, which reduces to the following result of R. Rado [**7**, Theorem 4.3], generalizing the Hoffman–Kuhn result to the case where X, Y may be infinite, when one puts $X' = X$:

COROLLARY 3. *Let $F = \{Y_x | x \in X\}$ be a family of finite subsets of a set Y. Let $Y' \subseteq Y$ be such that no element of Y' belongs to infinitely many sets Y_x. A necessary and sufficient condition that F have an SDR containing Y' is that both F satisfy condition H and that every finite subset L of Y' intersects at least $|L|$ of the sets in F.*

When $X' = X$, we have $F' = F$ in Corollary 2. The reader will recognize that the hypotheses of Corollary 3 imply that each set X_y is finite when $y \in Y'$, and that the condition on subsets L implies that $G' = \{X_y | y \in Y'\}$ satisfies the condition H.

The results above depend on Theorem 1 (whose easy proof was given), and on the result of Marshall Hall. One proof of the latter involves a rather natural application of Zorn's Lemma to the well-known completely finite version due to Philip Hall [**4**]. A second proof which is quite readable is given by Marshall Hall in his text *Combinatorial Theory* (Blaisdell, 1967, pp. 47–48). A third proof, for those familiar with logic, could apply a compactness argument to the result of Philip Hall for finitely many finite sets.

References

1. S. Banach, Un théorème sur les transformations biunivoques, Fund. Math., 6 (1924) 236–239.
2. G. Birkhoff and S. Mac Lane, A Survey of Modern Algebra, Macmillan, New York, 1953, pp. 362–363.
3. M. Hall, Jr., Distinct representatives of subsets, Bull. Amer. Math. Soc., 54 (1948) 922–926.

4. P. Hall, On representatives of subsets, J. London Math. Soc., 10 (1935) 26–30.
5. A. J. Hoffman and H. W. Kuhn, Systems of distinct representatives and linear programming, this MONTHLY, 63 (1956) 455–460.
6. N. S. Mendelsohn and A. L. Dulmage, Some generalizations of the problem of distinct representatives, Canad. J. Math., 10 (1958) 230–241.
7. L. Mirsky and H. Perfect, Systems of representatives, J. Math. Anal. Appl., 15 (1966) 520–568.
8. O. Ore, Theory of Graphs, AMS Colloq. Publ.XXXVIII (1962), 113 (Theorem 7.4.1).
9. H. Perfect and J. S. Pym, An extension of Banach's mapping theorem with applications to problems concerning representatives, Proc. Cambridge Phil. Soc., 62 (1966) 187–192. (See also the Addendum, Ibid. 63 (1967) 21.).

A NOTE ON LIOUVILLE'S THEOREM

A. LENARD
Department of Mathematics, Indiana University, Bloomington, IN 47405

A bounded entire function is a constant. This important theorem of Liouville is proved rather late in traditional courses on complex variables, and usually by making use of contour integration via Cauchy's Theorem and its developments. It may perhaps be of some pedagogical interest to have a maximally economical proof, in the sense that in it no concepts are used other than are needed for the statement and understanding of the theorem. It is my purpose to outline such a proof. Nothing more is required than the algebra of complex numbers and the theory of convergence in the complex domain, including the notion of absolute convergence, with the permissibility of multiplying absolutely convergent series term by term and rearranging such series. No notion of Calculus enters, not even differentiation or integration of real functions.

From this point of view an entire function is the sum

$$f(z)=\sum_0^\infty a_n z^n$$

of an everywhere converging, hence absolutely converging, power series. Assuming

$$|f(z)|\leqslant M$$

for all z, we need to show $a_n=0$ for all n except possibly for $n=0$. For this it is sufficient to show that

$$A(r)=\sum_0^\infty |a_n|^2 r^{2n}\leqslant M^2+2$$

for all positive r. This is accomplished as follows.

For any positive r choose $k=k(r)$ as the least positive integer such that

$$B(r)=\sum_0^\infty |a_n|r^n\sum_{m\geqslant k}|a_m|r^m\leqslant 1.$$

Then compute the arithmetical mean $F(r)$ of the k values of the function $|f|^2=f\bar{f}$ at the points $r\omega^j$ $(0\leqslant j\leqslant k-1)$, where $\omega=\exp(2\pi i/k)$. Since

$$\frac{1}{k}\sum_{j=0}^{k-1}\omega^{(n-m)j}=\begin{cases}1 & \text{if } n\equiv m \pmod{k},\\ 0 & \text{otherwise,}\end{cases}$$

the result is

$$F(r)=\sum\sum_{n\equiv m \pmod{k}} a_n\bar{a}_m r^{n+m},$$

and by hypothesis $F(r)\leqslant M^2$. Write $F(r)=A(r)+G(r)$, the first term containing all sum-

mands with $n = m$. Since $n \equiv m \pmod{k}$ and $n \neq m$ imply at least one of the inequalities $n \geqslant k$, $m \geqslant k$, and therefore $|G(r)| \leqslant 2B(r)$, one obtains

$$A(r) = F(r) - G(r) \leqslant F(r) + |G(r)| \leqslant M^2 + 2,$$

as required.

A few remarks are in order. The elementary nature and simplicity of this proof is purchased, of course, at the price of using the Weierstrassian power series definition of analyticity rather than the complex differentiability definition in the tradition of Cauchy-Riemann-Goursat. But that is the point: Complex numbers, convergence, and power series could be taught independently and *before* Calculus, albeit reversing present-day educational practice. Secondly, the number 2 on the right-hand side of the inequality for $A(r)$ plays no special role; it may evidently be replaced by any positive number no matter how small. Thus our proof actually yields $A(r) \leqslant M^2$; which is the best possible if one takes $M = \sup|f(z)|$. Finally, the above proof is closely related to, and is indeed the "discretized" version of, a well-known proof in which one uses averaging of $|f|^2$ with respect to arclength measure on the circle $|z| = r$. This yields directly

$$\frac{1}{2\pi}\int_0^{2\pi}|f(re^{i\theta})|^2 d\theta = A(r),$$

whence $A(r) \leqslant M^2$ is evident. The idea of replacing the integral by the finite Riemann approximating sum and thereby eliminating the last vestige of Calculus from the proof may have some value in other contexts where the needed integration theory is sufficiently cumbersome so that a more elementary procedure using sums instead of integrals has appeal.

This note is dedicated to the memory of Sándor Lénárd, uncle, engineer, and connoisseur of mathematics from whom, many years ago, the writer first heard about complex numbers.

A SELF-DUAL CARD TRICK BASED ON CONGRUENCES AND k-SHUFFLES

Kurt Eisemann
Department of Mathematics, San Diego State University, San Diego, CA 92182

Introduction. A card trick based on number-theoretic properties, resurrected by Martin Gardner [3] and analyzed in [4], has a sequel. The originator [1] has devised a second card trick [2], likewise based on congruences and k-shuffles, that exhibits the even more astonishing and extremely rare property of self-duality. The present paper describes in simplified form a generalized version of it, together with its mathematical basis. While the first trick dealt with two decks possessing mutual duality, the second is more impressive because it deals with a single deck that exhibits self-duality. It is less flexible than the first because rather than being valid for a variety of configurations, the given procedure applied to a chosen size deck furnishes essentially only a single array possessing this amazing property.

As a by-product, the notation and set-up described herein permit a further extension of the purely mathematical first card trick that significantly enhances the complexity of its aspect and its fascination.

Notation. The manner of enumerating card sequences, dealing, and the definition and execution of k-shuffles will be exactly as described in [4]. From a deck of playing cards, select cards numbered 1 (Ace) through b from each of s suits. Define the corresponding modulus as $m = bs + 1$, which we require to be prime. For illustration we shall take $b = 7$, $s = 4$, $m = 29$. Denote diamond, spade, heart and club by D_0, S_1, H_2, C_3 respectively—a redundant but mnemonic and effective notation. The 5 of hearts, for example, will be denoted $H_2 5$. This is to be interpreted as the positional notation for a 2-digit number in base b, with a *numerical value* in the decimal system of $2b + 5$ ($= 19$ here). The subscript of the letter designating the suit is the coefficient of b for computing the numerical value of a card. The shapes of the conventional

CENTER SECTION

(Vol. 93, No. 3, March 1986)

Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

T:	Textbook	13-18:	Grade Level
S:	Supplementary Reading	1-4 :	Time in Semesters
P:	Professional Reading	** :	Special Emphasis
L:	Undergraduate Library	?? :	Questionable

General, T(13-16), S*, L*. The Green Book: 100 Practice Problems for Undergraduate Mathematics Competitions. Kenneth Hardy, Kenneth S. Williams. Integer Pr, 1985, ix + 173 pp, $12 (P). [ISBN: 0-9692193-0-X] An inexpensive, but invaluable, collection of practice problems for students training for the William Lowell Putnam Mathematical Competition. Problems come with hints and complete solutions and cover a broad range of essential ideas and modes of thinking. LCL

General, L. Chess Variations: Ancient, Regional, and Modern. John Gollon. Charles E Tuttle, 1985, 233 pp, $6.25 (P). [ISBN: 0-8048-1122-9] Includes more than 40 chess variations with sample games of each. Many require special boards or pieces, but the author presents guidelines for their construction. Some of these chess variations are regarded in certain parts of the world as superior to Western chess. LCL

General, T(15-17), S, P, L. Introduction to Fuzzy Arithmetic: Theory and Applications. Arnold Kaufmann, Madan M. Gupta. Van Nostrand Reinhold, 1985, xvii + 351 pp, $44.95. [ISBN: 0-442-23007-9] Fuzzy sets, fuzzy arithmetic, and fuzzy functions are technical constructs designed to help deal with uncertainty in a computationally effective way. This volume is devoted exclusively to presenting a self-contained theory of fuzzy numbers and its applications. LCL

Precalculus, T(13-14: 4). Basic Technical Mathematics, Fourth Edition. Allyn J. Washington. Benjamin/Cummings, 1985, xvii + 749 pp, $31.95 [ISBN: 0-8053-9550-4]; Basic Technical Mathematics with Calculus, Fourth Edition, xix + 1068 pp, $33.95. [ISBN: 0-8053-9541-5] This edition has been largely rewritten but still incorporates the basic features of previous editions. As before, the second text includes the content of the first. (For Second Edition, see TR, October 1970, and August-September 1970.) New to this edition are calculator material, including an appendix with BASIC programs, sections on Newton's method and Simpson's rule and additional figures, exercises and examples. JNC

Precalculus, S(13). Basic Algebra and Trigonometry: A Traditional Approach. Pennsylvania State University (Dept. of Math.). Kendall/Hunt, 1985, ii + 138 pp, $9.95 (P). [ISBN: 0-8403-3703-5] A workbook for students (e.g., those taking beginning calculus courses) who need a supplementary review of traditional topics from high school algebra and trigonometry. Includes a section on determinants. LCL

Precalculus, T(13: 1-3). Basic Technical Mathematics. Stuart R. Porter, John F. Ernst. Addison-Wesley, 1985, xvi + 861 pp [ISBN: 0-201-05586-4]; Basic Technical Mathematics with Calculus, 1985, xv + 1227 pp [ISBN: 0-201-05589-9]; Preliminary Sample of Instructor's Manual. Presumes elementary algebra and plane geometry, but even reviews those. Includes usual topics of precalculus; further, one chapter introduces elementary BASIC. Problems of many levels of difficulty. Larger book is the smaller one plus six chapters on calculus (differentiation and integration and applications), and one on elementary probability and descriptive statistics. Instructor's Manual consists mainly of all answers not given in the texts themselves. DFA

Education, P*, L. Academic Preparation in Mathematics: Teaching for Transition From High School to College. Jeremy Kilpatrick. The College Board, 1985, vi + 86 pp, (P). [ISBN: 0-87447-222-9] A report from The College Board on goals for high school mathematics with suggested teaching strategies for achieving these goals. Following the 1983 College Board report "Academic Preparation for College," this special report on mathematics discusses especially the contribution that high school mathematics makes to the "basic academic competencies" of reading, writing, speaking, listening, reasoning, and studying. Since mathematics itself is also one of the basic competencies, the role of high school mathematics in successful preparation for college is stressed as especially important. LAS

Education, P. Mathematics, Science, and Technology Education: A Research Agenda. National Academy Pr, 1985, viii + 92 pp, (P). A report from the National Research Council to the National Institute

of Education assessing the current status of research on reasoning, instruction, settings, and new learning systems (i.e., technology). Principle recommendation: "A research agenda focused on the amount of time devoted to active teaching and learning of reasoning skills," also known as "quality learning time." The report itself is primarily a careful, well-organized summary of current research reports listed in an extensive bibliography. LAS

Logic, T(17-18: 1), P, L. Set Theory: An Introduction to Independence Proofs, Second Printing. Kenneth Kunen. Stud. in Logic & the Found. of Math., V. 102. Elsevier Science, 1983, xvi + 313 pp, $27.50 (P). [ISBN: 0-444-86839-9] Paperback reprint of a text published in 1980 (TR, June 1981). KS

Logic, T(17-18: 1), P. Topoi: The Categorical Analysis of Logic, Revised Edition. Robert Goldblatt. Stud. in Logic & the Found. of Math., V. 98. Elsevier Science, 1984, xvi + 551 pp, $69.25. [ISBN: 0-444-86711-2] Revision of a 1979 text (see TR, March 1980). Includes corrections plus new chapter on theory of geometric morphisms between Grothendieck topoi and model theoretic rendering of this theory due to Makkai and Reyes. KS

Logic, S(18), P*, L. Model-Theoretic Logics. Ed: J. Barwise, S. Feferman. Perspectives in Math. Logic. Springer-Verlag, 1985, xviii + 893 pp, $140. [ISBN: 0-387-90936-2] A major piece of work which makes available an invaluable resource for anyone hoping to reach the frontiers of research in abstract model theory. The individual contributions provide an introduction to the main ideas, examples, and results of the literature, and presents a picture of a rapidly evolving subject. Concludes with a unified bibliography containing 1,261 entries! LCL

Logic, P. Logic Colloquium '82. Ed: G. Lolli, G. Longo, A. Marcja. Stud. in Logic & the Found. of Math., V. 112. Elsevier Science, 1984, viii + 358 pp, $52. [ISBN: 0-444-86876-3] Proceedings of a conference held in Florence, Italy, August 23-28, 1982. Invited research papers on model theory, categorical logic, lambda calculus, connections between logic and theoretical computer science. Survey by Smoryński on nonstandard models of arithmetic. KS

Graph Theory, P. Lecture Notes in Mathematics-1073: Graph Theory, Singapore 1983. Ed: K.M. Koh, H.P. Yap. Springer-Verlag, 1984, xiii + 335 pp, $16 (P). [ISBN: 0-387-13368-2] Proceedings of First Southeast Asian Graph Theory Colloquium held May 10-28, 1983. Thirty research papers plus list of open problems. KS

Graph Theory, T(13-14: 1), S, L. Introductory Graph Theory. Gary Chartrand. Dover, 1985, xii + 294 pp, $5.95 (P). [ISBN: 0-486-24775-9] Corrected reprint of Graphs as Mathematical Models published by Prindle, Weber and Schmidt in 1977 (see TR, March 1978). KS

Graph Theory. Ten Applications of Graph Theory. Hansjoachim Walther. Math. & Its Applic. D Reidel, 1984, xii + 252 pp, $39.50. [ISBN: 90-277-1599-8] A translation of Anwendung der Graphentheorie published by Deutscher Verlag der Wissenschaften in 1978, containing ten extensively treated applications of graph theory to operations research, primarily. SS

Combinatorics, P. Polyhedral Combinatorics and the Acyclic Subdigraph Problem. M. Jünger. Res. & Expos. in Math., V. 7. Heldermann Verlag, 1985, x + 128 pp, $36 (P). [ISBN: 3-88538-207-5] A volume on polyhedral combinatorics, apparently for the researcher, but actually accessible to a wider audience of readers having only a modest background in combinatorics. Attention is given to optimization and computational problems as well as to the basic theory. SS

Discrete Mathematics, T(16-17: 1, 2), L. Combinatorics for Computer Science. S. Gill Williamson. Comput. & Math. Ser. Computer Science Pr, 1985, xliii + 479 pp, $39.95. [ISBN: 0-88175-020-4] The two parts of this book treat enumeration and graph theory. Topics covered are those of especial interest to computer scientists and applied mathematicians. Organized for use in a seminar setting. AO

Linear Algebra, T(16-17: 1, 2), L. The Theory of Matrices, Second Edition with Applications. Peter Lancaster, Miron Tismenetsky. Comput. Sci. & Appl. Math. Academic Pr, 1985, xv + 570 pp, $59. [ISBN: 0-12-435560-9] An extensive revision of the previous edition (TR, October 1969). The first part is a fairly standard treatment of matrices and linear algebra. The remainder develops several topics of importance in applications: variational methods, perturbation theory, generalized inverses, stability theory, nonnegative matrices, etc. AO

Algebra, T*(16-17: 1), S, L*. Applied Abstract Algebra. Rudolf Lidl, Günter Pilz. Undergrad. Texts in Math. Springer-Verlag, 1984, xviii + 545 pp, $39. [ISBN: 0-387-96035-X] Readable text for applied algebra course which assumes first course in abstract algebra and linear algebra. Covers applications of Boolean algebras, finite fields and semigroups including switching circuits, coding theory and automata. Extensive problem sets at ends of sections with many computational problems. KS

Algebra, T(15-16), S, L. Applications of Abstract Algebra. George Mackiw. Wiley, 1985, 184 pp, $11.95 (P). [ISBN: 0-471-81078-9] A collection of applications of abstract algebra (including exact computing, error correcting codes, block designs, cryptography, crystallography) which can be used as a supplement in a beginning course in abstract algebra, or as a separate course in applications. It does not include more advanced applications such as those to quantum mechanics and advanced physics. LCL

Calculus, T(13: 2). Mathematics and Calculus with Applications, Second Edition. Margaret L. Lial, Charles D. Miller. Scott Foresman, 1985, xxxvi + 924 pp, $28.95. [ISBN: 0-673-15896-9] Informal, intuitive, non-rigorous presentation of finite mathematics, mathematics of finance, elementary calculus, discrete and continuous probability. Applications are chiefly to business and economics; most chapters include documented 2-3 page "extended applications." (First Edition, TR, December 1980.) PZ

Calculus, T, S(13). Student's Guide to Calculus by J. Marsden and A. Weinstein, Volume I. Frederick H. Soon. Springer-Verlag, 1985, 312 pp, $14.95 (P). [ISBN: 0-387-96207-7] Contains a prerequisite quiz (with answers), goals, study hints, solutions to every-other odd exercise and a section quiz (with answers) for each section of the text (see TR, May 1980). JNC

Calculus, S(13). Student's Solutions Manual to Accompany Simmons: Calculus with Analytic Geometry. Anthony Barcellos, Dean Hickerson. McGraw-Hill, 1985, v + 629 pp, $19.95 (P). [ISBN: 0-07-057529-0] Detailed solutions to all odd-numbered problems. LCL

Calculus, T(13: 1). Calculus with Applications, Third Edition. Margaret L. Lial, Charles D. Miller. Scott Foresman, 1985, xxxvi + 651 pp, $27.95. [ISBN: 0-673-15895-0] Terse, "intuitive and visual" presentation essentially without motivation or proof; includes chapters on multivariable calculus, differential equations, calculus of probability. Supposedly a complete revision of previous editions but "portions of this book appear in Mathematics and Calculus with Applications, Second Edition." JNC

Real Analysis, T(16-17: 1), P, L*. An Introduction to Nonstandard Real Analysis. Albert E. Hurd, Peter A. Loeb. Pure & Appl. Math., V. 118. Academic Pr, 1985, xii + 232 pp, $35. [ISBN: 0-12-362440-1] An introduction to nonstandard analysis and some of its applications in elementary calculus, topology, and integration theory. No prior background in mathematical logic is assumed. AO

Real Analysis, P, L*. The Banach-Tarski Paradox. Stan Wagon. Ency. of Math. & Its Applic. Cambridge U Pr, 1985, xvi + 251 pp, $37.50. [ISBN: 0-521-30244-7] Begins with an elementary discussion of the classical measure theoretic paradox that any sphere in R^3 (but not in the line or the plane) can be partitioned into finitely many pieces which can be reassembled by rigid motions into a sphere of any other size. Links this paradox to related paradoxes of measure theory in low dimension and to recent (post-1980) results in the foundations of measure theory. LAS

Differential Equations, S(17-18), P. Problèmes aux Limites non Linéaires pour Certaines Classes D'Equations Différentielles Ordinaires. Ronald B. Guenther. Pr U Montreal, 1985, 162 pp, $19 (P). [ISBN: 2-7606-0697-X] An introduction to important results concerning existence and uniqueness of and numerical approximation to solutions of various two-point boundary value problems. From lectures at a seminar in topological methods in nonlinear analysis at Montreal in summer 1983. Translated from English. Accessible to non-specialists. DFA

Partial Differential Equations, P. Functional Integration and Partial Differential Equations. Mark Freidlin. Annals of Math. Stud., No. 109. Princeton U Pr, 1985, ix + 545 pp, $60; $19.95 (P). [ISBN: 0-691-08354-1; 0-691-08362-2] Utilizes probability measures to study differential equations. This approach simplies a number of problems and enables one to discover new effects. LCL

Partial Differential Equations, P. The Mathematics of Finite Elements and Applications V: MAFELAP 1984. Ed: J.R. Whiteman. Academic Pr, 1985, xviii + 650 pp, $59. [ISBN: 0-12-747255-X] Proceedings of the fifth conference on the mathematics of finite elements and applications held in May 1984 at Brunel University. Contains the texts of eleven invited papers, thirty-four contributed papers, and the abstracts of the poster papers. AO

Partial Differential Equations, P. Nonlinear Variational Problems. Ed: A. Marino, et al. Res. Notes in Math., V. 127. Pitman, 1985, 124 pp, $14.95 (P). [ISBN: 0-273-08670-7] Seven invited lectures and sixteen short communications presented to an international workshop held at Isola d'Elba, Italy in September 1983. DFA

Partial Differential Equations, P. Microdifferential Systems in the Complex Domain. Pierre Schapira. Grund. der math. Wissenschaften, B. 269. Springer-Verlag, 1985, x + 214 pp, $34.50. [ISBN: 0-387-13672-X] Introduction to and applications of the theory of systems of microdifferential equations, viewed as modules over rings of operators. First two chapters develop analytic and algebraic machinery, third applies it to various Cauchy problems. Four brief appendices: symplectic geometry, homological algebra, sheaves, results of analytic geometry. PZ

Numerical Analysis, T(15-16: 1, 2), L. Numerical Mathematics: Theory and Computer Applications. Carl-Erik Fröberg. Benjamin/Cummings, 1985, xi + 436 pp, $37.95. [ISBN: 0-8053-2530-1] An extensively revised and updated version of the Second Edition of the author's Introduction to Numerical Analysis. The general outline of the older book has been retained, but many new topics and exercises have been added. AO

Numerical Analysis, P. Lecture Notes in Mathematics-1129: Numerical Analysis, Lancaster 1984. Ed: P.R. Turner. Springer-Verlag, 1985, xiv + 179 pp, $12 (P). [ISBN: 0-387-15234-2] Lecture notes on four topics: algorithmic aspects of approximation theory, optimal recovery, finite element method, and variational theory for boundary value problems. BC

Numerical Analysis, P. Large-Scale Computations in Fluid Mechanics. Ed: Bjorn E. Engquist, Stanley Osher, Richard C.J. Somerville. Lect. in Appl. Math. V. 22. AMS, 1985, $110 set [ISBN: 0-8218-1122-3]. Part 1, xv + 370 pp; Part 2, x + 409 pp. Proceedings of the Fifteenth AMS-SIAM Summer Seminar on Applied Mathematics held at the Scripps Institution of Oceanography from June 27 to July 8, 1983. AO

Numerical Analysis, P. The Mathematical Basis of Finite Element Methods with Applications to Partial Differential Equations. Ed: David F. Griffiths. New Ser., V. 2. Clarendon Pr, 1984, 189 pp, $24.95. [ISBN: 0-19-853605-4] Finite element methods are numerical techniques motivated by elliptic partial differential equations arising in structural mechanics (and elsewhere), and based on classical variational principles. These nine papers were given at an expository conference at Imperial College in January, 1983. BC

Numerical Analysis, P. Computational Mathematics. Ed: Andrzej Wakulicz. Banach Center Pub., V. 13. PWN, 1984, 792 pp. [ISBN: 83-01-04968-5] Fifty papers (fourteen in Russian) on numerical methods in differential equations and linear algebra. BC

Functional Analysis, S(18), P. An F-space Sampler. N.J. Kalton, N.T. Peck, James W. Roberts. London Math. Soc. Lecture Note Ser., V. 89. Cambridge U Pr, 1984, xii + 240 pp, $24.95 (P). [ISBN: 0-521-27585-7] These notes present some of the principle results from the theory of complete metric linear spaces which are not necessarily locally convex. (Such spaces in which the Hahn-Banach theorem holds are locally convex.) F-space examples include l_p, L_p, and H^p where $0 < p < 1$. Other important sections cover the Hahn-Banach extension property, lifting theorems, and transitivity and small operators. MU

Functional Analysis, T(15-17: 1, 2), S, L. Introduction to Integral Equations with Applications. Abdul J. Jerri. Pure & Appl. Math., V. 93. Dekker, 1985, x + 254 pp, $39.75. [ISBN: 0-8247-7293-8] Leisurely introduction to Volterra and Fredholm integral equations, assuming only one course in differential equations. Motivated from page one by mathematical models, especially from biology and mechanics. Concluding chapter on existence theorems. Exercises and answers. BC

Functional Analysis, P. Dual Algebras with Applications to Invariant Subspaces and Dilation Theory. Hari Bercovici, Ciprian Foiaş, Carl Pearcy. CBMS Reg. Conf. Ser. in Math., No. 56. AMS, 1985, xi + 108 pp, $16 (P). [ISBN: 0-8218-0706-4] Expanded lecture notes from a 1984 NSF/CBMS regional conference at Tempe, Arizona. The objects of study are certain subalgebras of bounded linear operators on a separable Hilbert space; they are used to obtain results on invariant subspaces and dilation theory for Hilbert space operators. PZ

Functional Analysis, P. Theory of Multipliers in Spaces of Differentiable Functions. V.G. Maz'ya, T.O. Shaposhnikova. Mono. & Stud. in Math., V. 23. Pitman, 1985, xiii + 344 pp, $49.95. [ISBN: 0-273-08638-3] A description of multipliers, their properties and applications, for students and researchers interested in function spaces and their relation to partial differential equations and operator theory. The theory of Fourier multipliers is not covered. Assumes a familiarity with Sobolev space theory. LCL

Analysis, T(17-18: 1, 2), S*, P. Harmonic Analysis on Symmetric Spaces and Applications, I. Audrey Terras. Springer-Verlag, 1985, xv + 341 pp, $39 (P). [ISBN: 0-387-96159-3] Harmonic analysis in non-Euclidean space presented as a natural extension of familiar Fourier analysis on R^n. (A table compares the main formulas from each.) Applications galore, spanning everything from algebraic number theory to microwave engineering. Many, many exercises. Extensive bibliography. Well written--fascinating mathematics on every page. BC

Topology, T*(14-16: 1), S*, P, L.** The Shape of Space: How to Visualize Surfaces and Three-Dimensional Manifolds. Jeffrey R. Weeks. Pure & Appl. Math., V. 96. Dekker, 1985, x + 324 pp, $49.75. [ISBN: 0-8247-7437-X] A well-written, superbly illustrated informal introduction to geometric topology of two and three dimensions, ideally suited for bright high school students or nascent undergraduate mathematics majors. Begins with gluing surfaces and three-manifolds; ends with a survey of Thurston's work on geometries on three-manifolds with applications to models of the universe. Includes exercises (with answers) and a bibliography. LAS

Operations Research, S*(15-17), L*. Model Building in Mathematical Programming, Second Edition. H.P. Williams. Wiley, 1985, xv + 349 pp, $19.95 (P). [ISBN: 0-471-90606-9] Discusses the general principles of model building in mathematical programming. Also presents twenty problems in which these principles can be applied together with suggested models and computational results. (First Edition, TR, October 1978.) AO

Operations Research, T(14-15: 1), S, L. Mathematics in the Social and Life Sciences: Theories, Models and Methods. M.A. Ball. Math. & Its Applic. Halsted Pr, 1985, 296 pp, $57. [ISBN: 0-470-20191-6] The aim of the book is to show how mathematical models and theories are developed in the social and life sciences. The book develops three self-contained subject areas: microeconomics, games (conflict and cooperation), and population dynamics. Models are developed carefully with attention to assumptions and credibility. Note price! LCL

Operations Research, T(14-16: 1), S, L. Decision Making, Models and Algorithms: A First Course. Saul I. Gass. Wiley, 1985, xvi + 412 pp, $32.95. [ISBN: 0-471-80963-2] A powerful textbook divided into five parts (general, linear programming (2), networks, games). Each part ends with a substantial chapter that combines further discussions, extensions, and exercises which challenge students

to active involvement in the "decision framework" (model formulation, algorithm development, variations under changing assumptions, solution techniques). LCL

Optimization, P. Lecture Notes in Economics and Mathematical Systems-243: Integer Programming and Related Areas: A Classified Bibliography, 1981-1984. Ed: R. von Randow. Springer-Verlag, 1985, xx + 386 pp, $29.20 (P). [ISBN: 0-387-15226-1] The fourth in a series of bibliographies on this topic, this volume covers the period from mid-1981 to autumn 1984. AO

Optimization, T(15-16: 1), L. Linear and Nonlinear Programming: An Introduction to Linear Methods in Mathematical Programming. Roger Hartley. Ser. in Math. & Its Applic. Halsted Pr, 1985, 221 pp, $39.95; $16.95 (P). [ISBN: 0-470-20178-9] An introductory textbook covering linear programming and closely related topics: network problems, multiple objectives, integer programming and quadratic programming. An informal but rigorous presentation presupposing only a minimal mathematical background. AO

Optimization, P. Numerical Optimization Techniques. Yurij G. Evtushenko. Transl. Ser. in Math. & Engin. Optimization Software, 1985, xiv + 558 pp, $68. [ISBN: 0-911575-07-3] Describes several techniques for the numerical solution of nonlinear programming problems and applications of these methods, especially in the solution of optimal control problems for ordinary differential equations. AO

Optimization, P. Computational Mathematical Programming. Ed: Klaus Schittkowski. NATO ASI Ser. F, V. 15. Springer-Verlag, 1985, 451 pp, $50.20. [ISBN: 0-387-15180-X] Written versions of the main lectures presented at the NATO Advanced Study Institute held in Bad Windsheim, Germany from July 23 to August 2, 1984. The papers in this volume are tutorial in nature. AO

Dynamical Systems, P. Chaos, Fractals, and Dynamics. Ed: P. Fischer, William R. Smith. Lect. Notes in Pure & Appl. Math., V. 98. Dekker, 1985, viii + 261 pp, $71.50 (P). [ISBN: 0-8247-7325-X] Eighteen papers presented at or resulting from two conferences at the University of Guelph in March 1981 and 1983. Applied and multidisciplinary (physics, chemistry, biology, engineering). Includes many illustrations, some computer-generated. DFA

Dynamical Systems, P. Symmetries for Dynamical and Hamiltonian Systems. H.M.M. ten Eikelder. CWI Tract, V. 17. Math Centrum, 1985, ii + 191 pp, Dfl. 27,40 (P). [ISBN: 90-6196-288-9] The first half is devoted to the general theory for systems on an arbitrary manifold; the second half to applications. LCL

Control Theory, P. Computational Methods for Optimizing Distributed Systems. K.L. Teo, Z.S. Wu. Math. in Sci. & Eng., V. 173. Academic Pr, 1984, xiii + 317 pp, $62. [ISBN: 0-12-685480-7] An in-depth treatment of optimal control problems involving linear second-order parabolic partial differential equations. Computational techniques for these problems are developed using a first-order strong variational method. AO

Probability, T*(14-15: 1). Probability and Statistical Inference, Volume 1: Probability, Second Edition. J.G. Kalbfleisch. Texts in Stat. Springer-Verlag, 1985, xiii + 343 pp, $29.80. [ISBN: 0-387-96144-5] A revision of the 1979 First Edition (TR, March 1980). Minor changes and additional exercises. An attractive option for a one-semester course in probability or the first half of a one-year probability/statistics sequence. TAV

Statistics, T, P. Lecture Notes in Statistics-28: Differential-Geometrical Methods in Statistics. Shun-ichi Amari. Springer-Verlag, 1985, v + 290 pp, $19.60 (P). [ISBN: 0-387-96056-2] Inspired by C.R. Rao, Efron and others, Amari's goal is to construct a differential geometric theory of statistics. Covers introductory differential geometry, geometric properties of families of distributions, differential geometry of statistical manifolds, asymptotic theory of estimation and testing in the framework of curved exponential families. KK

Statistics, T(16-18: 1), S, P. Sequential Analysis: Tests and Confidence Intervals. David Siegmund. Ser. in Stat. Springer-Verlag, 1985, xi + 272 pp, $32. [ISBN: 0-387-96134-8] Recent developments in sequential hypotheses testing on a non-Bayesian, non-decision-theoretic context. The early chapters emphasize statistical ideas, and difficult proofs are presented later on. FLW

Statistics, T(17-18: 1-3), S. Introduction to Probability and Mathematical Statistics. Václav Fabian, James Hannan. Wiley, 1985, xi + 466 pp, $42.95. [ISBN: 0-471-25023-6] A unified mathematical treatment of probability and statistics at the measure theoretic level. FLW

Statistics, S(15-18), P. The Frontiers of Modern Statistical Inference Procedures. Ed: Edward J. Dudewicz. American Sciences Pr, 1985, xi + 493 pp, $79.95 (P). [ISBN: 0-935950-07-9] Proceedings of the July 1982 conference held in Hawaii on inference procedures associated with statistical ranking and selection. FLW

Statistics, T(13: 1). Elementary Statistics. Donna H. Skane. Addison-Wesley, 1985, xviii + 414 pp. [ISBN: 0-201-06751-X] Presupposes only high school algebra. The usual topics. There is a software package for the Apple II microcomputer that can be used with the text. FLW

Statistics, L.** Applied Statistics Algorithms. Ed: P. Griffiths, I.D. Hill. Ser. in Math. & Its Applic. Halsted Pr, 1985, 307 pp, $54.95. [ISBN: 0-470-20184-3] A collection of thirty-two algorithms from the journal Applied Statistics. All algorithms are given in standard Fortran and have

been updated to follow the journal's current guidelines and to incorporate corrections. Remarks accompanying each algorithm detail changes made to the originally published code. AO

Statistics, T(16-18: 1). Multiple Regression and the Analysis of Variance and Covariance, Second Edition. Allen L. Edwards. WH Freeman, 1985, xv + 221 pp, $12.95 (P); $19.95. [ISBN: 0-7167-1704-2; 0-7167-1703-4] Concerned with the relationship between multiple regression and the analysis of variance and covariance. Intended as a text to supplement or follow a course on analysis of variance for students in the behavioral sciences. The exercises require only hand calculators. (First Edition, TR, May 1980.) FLW

Statistics, S(17-18), P. Some Large Deviation Results in Statistics. A.D.M. Kester. CWI Tract, V. 18. Math Centrum, 1985, iii + 135 pp, Dfl. 19,10 (P). [ISBN: 90-6196-289-7] Considers probabilities of gross errors in point estimators and the Bahadur efficiency of two-sample conditional tests in exponential families. FLW

Statistics, T(15-18: 1, 2), S, L**.** Exploring Data Tables, Trends, and Shapes. Ed: David C. Hoaglin, Frederick Mosteller, John W. Tukey. Wiley, 1985, xxii + 527 pp, $39.95. [ISBN: 0-471-09776-4] A companion text for Understanding Robust and Exploratory Data Analysis (TR, December 1983) by the same authors. An important presentation of advanced robust and exploratory techniques. FLW

Statistics, T(13: 1). Essential Statistics. D.G. Rees. Chapman & Hall, 1985, xiii + 234 pp, $15.95 (P). [ISBN: 0-412-26440-4] A non-mathematical treatment with the usual topics plus the Poisson distribution, some non-parametric tests and subjective probabilities. The problems presuppose use of a hand-held calculator. FLW

Computer Programming, P. Foundations of Programming. Jacques Arsac. Transl: Fraser Duncan. APIC Stud. in Data Proc., V. 23. Academic Pr, 1985, xxxii + 265 pp, $29.50. [ISBN: 0-12-064460-6] Presents a programming methodology based on recurrence, recursion, and iteration. Many examples are given to illustrate the techniques advocated by the author. AO

Computer Programming. Beyond Mindstorms: Teaching with IBM LOGO. Joyce Tobias, et al. Holt, Rinehart & Winston, 1985, xiii + 431 pp, $21.95 (P). [ISBN: 0-03-071722-1] The title should read "Teaching IBM LOGO." The references to non-computer programming curriculum are sparse and sometimes contrived: this is really an elementary programming text. MW

Computer Programming, P*, L. Fortran Optimization, Revised Edition. Michael Metcalf. APIC Stud. in Data Proc., V. 25. Academic Pr, 1985, xii + 253 pp, $26. [ISBN: 0-12-492482-4] A detailed survey of techniques that can be used to increase the run-time efficiency of Fortran programs. Appendices describe ANSI Fortran 77 and provide a layout program. Recommended for serious Fortran programmers. (First Edition, TR, January 1985.) AO

Computer Programming, T(15-16: 1), P. Understanding Ada: A Software Engineering Approach. Gary Bray, David Pokrass. Wiley, 1985, xv + 352 pp, $19.95 (P). [ISBN: 0-471-87833-2] An informal introduction to the Ada programming language emphasizing language features that support good software engineering practice. Assumes the reader is familiar with high-level language programming and the software development process. AO

Software Systems, P. User Interface Management Systems. Ed: Günther E. Pfaff. Eurographic Seminars. Springer-Verlag, 1985, xii + 224 pp, $34.50. [ISBN: 0-387-13803-X] Proceedings of a 1983 workshop in Germany on UIMS (systems to manage dialogue between users and interactive graphics sytems). Reports and individual papers on four areas: role, model, structure, and construction of UIMS; dialogue specification tools; interfaces and implementation; user's conceptual model. RM

Software Systems, S(16-17), P, L. Synchronization of Parallel Programs. F. André, D. Herman, J.-P. Verjus. Transl: J. Howlett. Ser. in Sci. Comput. MIT Pr, 1985, x + 110 pp, $20. [ISBN: 0-262-01085-2] Introduces a method for expressing cooperation and competition problems together with different synchronization and communication methods for implementing them. Discusses both centralized and distributed approaches to the problem of controlling concurrency. AO

Computer Science, T(15-16: 1), P, L*. Prolog for Programmers. Feliks Kluźniak, Stanisław Szpakowicz. APIC Stud. in Data Proc., V. 24. Academic Pr, 1985, xii + 400 pp, $47.50. [ISBN: 0-12-416520-6] An introduction to the Prolog programming language with examples of its use and implementation. Written for programmers with substantial experience with procedural languages. AO

Computer Science, S(16-18), P, L. Lecture Notes in Computer Science-190: Distributed Systems. M.W. Alford, et al. Springer-Verlag, 1985, vi + 573 pp, $31.50 (P). [ISBN: 0-387-15216-4] Tutorial papers from the Advanced Course on Distributed Systems--Methods and Tools for Specification held in Munich, Germany in April 1984 and again in April 1985. AO

Computer Science, S(16-18), P, L. Lecture Notes in Computer Science-184: Local Area Networks: An Advanced Course. Ed: D. Hutchison, J.A. Mariani, W.D. Shepherd. Springer-Verlag, 1985, viii + 497 pp, $31 (P). [ISBN: 0-387-15191-5] Lecture notes prepared for the Advanced Course on Local Area Networks held at Strathclyde University in July 1983. Topics covered include technologies, protocols, distributed computing, and network operating systems. AO

Computer Science, T(16-17: 1, 2), L*. Computational Geometry, An Introduction. Franco P. Preparata, Michael Ian Shamos. Texts & Mono. in Comp. Sci. Springer-Verlag, 1985, xii + 390 pp, $45. [ISBN:

0-387-96131-3] Computational geometry is a relatively new field concerned with the study of geometrical algorithms. This book is a comprehensive introduction to the subject emphasizing algorithm design and analysis of worst-case complexity. AO

Computer Science, P. Combinatorial Algorithms on Words. Ed: Alberto Apostolico, Zri Galil. NATO ASI Ser. F: Comp. & Sys. Sci., V. 12. Springer-Verlag, 1985, viii + 361 pp, $50.20. [ISBN: 0-387-15227-X] Twenty-four papers from a workshop held in June 1984 in Maratea, Italy. Provides an overview of recent work in the area. Emphasizes combinatorics and computational complexity. AO

Computer Science, S(15-17), L*. Operating Systems: Structures and Mechanisms. Philippe A. Janson. Academic Pr, 1985, xx + 267 pp, $29.50. [ISBN: 0-12-380230-X] An introduction to operating systems emphasizing the practical rather than the theoretical aspects of the subject. Could be used as a supplement to lectures or as a more theoretical textbook. AO

Computer Science, S(15-16), P, L. Lucid, The Dataflow Programming Language. William W. Wadge, Edward A. Ashcroft. APIC Stud. in Data Proc., V. 22. Academic Pr, 1985, xiii + 310 pp, $39.50. [ISBN: 0-12-729650-6] Lucid is a nonprocedural programming language based on a demand-driven dataflow model of computation that supports iteration. Provides an introduction to the philosophy of dataflow computing as well as to the language itself. AO

Computer Science, P. Lecture Notes in Computer Science-199: Fundamentals of Computation Theory. Ed: Lothar Budach. Springer-Verlag, 1985, xii + 542 pp, $27.40 (P). [ISBN: 0-387-15689-5] The proceedings of the 1985 International Conference on Fundamentals of Computation Theory held in Cottbus, GDR. AO

Applications, T(15-16: 2), S, P, L.** Introduction to Applied Mathematics. Gilbert Strang. Wellesley-Cambridge Pr (Box 157, Wellesley, MA 02181), 1986, xii + 758 pp, $39. [ISBN: 0-9614088-0-4] A dramatic, innovative text that integrates continuous and discrete methods (differential and matrix equations) with an algorithmic spirit linked to readily available scientific software packages. The guiding principle is the triple product A^TCA: from it flows both standard and innovative contemporary applications such as the Kalman filter, strange attractors, and Karmarkar's approach to linear programming, all within a contemporary framework of differential equations, equilibrium problems, and dynamical equations. A privately published labor of love, this text overflows with unifying insight: it makes its subject truly beautiful. LAS

Applications, S(16-17), P, L. Environmental and Natural Resource Mathematics. Ed: Robert W. McKelvey. Proc. of Symp. in Appl. Math., V. 32. AMS, 1985, xii + 143 pp, $28. [ISBN: 0-8218-0087-6] Papers by Graciela Chichilnisky, Colin Clark, Frank Clarke, Maureen Cropper, Geoffrey Heal and Richard Plant. From the AMS 1984 summer short course in Eugene, Oregon. The final paper is a transcript of a panel discussion by the authors on the role of mathematicians in natural resource modeling. DFA

Applications (Artificial Intelligence), P. Knowledge-based Interpretation of Outdoor Natural Color Scenes. Yuichi Ohta. Res. Notes in Artif. Intell., V. 4. Pitman, 1985, 136 pp, $19.50 (P). [ISBN: 0-273-08673-1] Rule-based region analyzer for outdoor color scenes. Uses top-down and bottom-up control, image partition into regions, symbolic management of regions, use of color information in segmentation, generalization of Karhunen-Loewe transform, to provide symbolic description of scene from its 2-D color image. RM

Applications (Artificial Intelligence), S(15-17), L. Learning to Solve Problems by Searching for Macro-Operators. Richard E. Korf. Research Notes in Artif. Intell., V. 5. Pitman, 1985, 147 pp, $19.50 (P). [ISBN: 0-273-08690-1] Learning efficient strategies for solving problems by searching for macro-operators (sequences of primitive operators) for problems (e.g., Rubik's cube) where subgoals are non-serializable (previously satisfied subgoals violated later in solution path) hence means-ends analysis, heuristic evaluation methods not useful. Operator decomposability, search for macros are now weak methods, paradigms for learning; utility of interleaving learning and problem solving. RM

Applications (Artificial Intelligence), S(17-18), P. Empirical Analysis for Expert Systems. Peter G. Politakis. Res. Notes in Artif. Intell., V. 6. Pitman, 1985, 187 pp, $19.50 (P). [ISBN: 0-273-08663-4] Description of SEEK, an interactive artificial intelligence system for developing expert systems (e.g., for rheumatic disease diagnosis). Expert knowledge and representative cases are used to integrate performance and verification concerns into development process. Expert's model assumed accurate; analysis of misdiagnosed cases suggests rule modification (generalization, specialization, change in confidence) to fine tune rules. RM

Applications (Artificial Intelligence), P. Qualitative Reasoning About Physical Systems. Ed: Daniel G. Bobrow. MIT Pr, 1985, 495 pp, $22.50 (P). [ISBN: 0-262-02218-4] A collection of eight papers describing current work in the area of qualitative reasoning. The papers are reprinted from Artificial Intelligence: An International Journal. AO

Applications (Computer Graphics), S(16-17), P. Computer Animation: Theory and Practice. Nadia Magnenat-Thalmann, Daniel Thalmann. Comp. Sci. Workbench. Springer-Verlag, 1985, xiii + 240 pp, $29.50. [ISBN: 0-387-70005-6] All aspects of computer animation--history, case studies (e.g., of the authors film "Dream Flight"), animation-oriented systems and languages, graphics editors. Key frame interpolation, painting, hidden surface removal, reflection, texture, etc., fractal techniques, human modeling. Existing films and systems discussed in detail; image synthesis partially explored.

Applications (Economics), P. Isolation and Aggregation in Economics. Ekkehart Schlicht. Springer-Verlag, 1985, xi + 112 pp, $22.50. [ISBN: 0-387-15254-7]

Applications (Economics), P. Physical Models and Equilibrium Methods in Programming and Economics. B.S. Razumikhin. Math. & Its Applic. D Reidel, 1984, xv + 351 pp, $64. [ISBN: 90-277-1644-7] Analogies between the principles and techniques of physics (analytical mechanics and thermodynamics) and mathematical economics are explored. Standard methods of mathematical programming and models from mathematical economics are given physical interpretations. AO

Applications (Engineering), T(15), P. Random Processes: A Mathematical Approach for Engineers. Robert M. Gray, Lee D. Davisson. Inform. & Syst. Sci. Ser. Prentice-Hall, 1986, xii + 305 pp, $32.95. [ISBN: 0-13-752882-5] Attempting a compromise between cookbook and abstract approaches, the authors present a mathematically-sound treatment with special emphasis on the input/output analysis of linear systems. Curiously, standard topics such as the Poisson process are only casually mentioned. TAV

Applications (Linguistics), P. Generalized Phrase Structure Grammar. Gerald Gazdar, et al. Harvard U Pr, 1985, xii + 276 pp, $35; $11.50 (P). [ISBN: 0-674-34455-3; 0-674-34456-1] Theory of generative grammar--syntactic features, feature instantiation principles, phrase structure and metarules, model theoretic semantics (denotation in a model) within Montague's possible worlds framework. Language properties are theorems, not axioms, with syntactic description in a single level system. Illustrated with GPSG analysis of English. RM

Applications (Physics), P. Particle Systems, Random Media and Large Deviations. Ed: Richard Durrett. Contemp. Math., V. 41. AMS, 1984, ix + 380 pp, $32 (P). [ISBN: 0-8218-5042-3] Proceedings of the AMS Summer Research Conference on the "Mathematics of Phase Transitions" held in June 1984. AO

Applications (Physics), P. The Recursion Method and Its Applications. Ed: D.G. Pettifor, D.L. Weaire. Ser. in Solid-State Sci., V. 58. Springer-Verlag, 1985, viii + 179 pp, $29. [ISBN: 0-387-15173-7] Proceedings of a conference held at Imperial College, London, in September 1984. The recursion method is a computational technique based on the Lanczos algorithm for tridiagonalization of matrices. AO

Applications (Physics), P. Open Quantum Systems and Feynman Integrals. Pavel Exner. Fund. Theories of Physics. D Reidel, 1985, xix + 356 pp, $64. [ISBN: 90-277-1678-1] An introduction to the general theory of unstable quantum systems using functional-analytic methods. Special attention is given to the extension of Feynman path integral techniques to certain non-selfadjoint second-order differential operators. AO

Applications (Physics), P. Entropy, Large Deviations, and Statistical Mechanics. Richard S. Ellis. Grund. der math. Wissenschaften, B. 271. Springer-Verlag, 1985, xiv + 364 pp, $54. [ISBN: 0-387-96052-X] Using entropy as a unifying concept, presents an introduction to the theory of equilibrium statistical mechanics and the large deviation properties of stochastic systems. AO

Applications (Physics), T(18), P.** Mathematical Theory of Non-linear Elasticity. A. Hanyga. Ser. in Math. & Its Applic. Halsted Pr, 1985, 432 pp, $80. [ISBN: 0-470-27493-X] While maintaining mathematical rigor, this book covers a very wide range of practical applications and provides a unified theory of the physical fields which occur in the description of physical phenomena with singularities. The main part of the book deals with the surface singularity, describing a regular and orientable surface moving on the continuum, and shows how the compatibility conditions for a function which suffers the singularity on the surface are derived. Written like a fine graduate-level mathematics text. MU

Applications (Psychology), T(17-18: 2), S, P. Elements of Psychophysical Theory. Jean-Claude Falmagne. Psych. Ser., No. 6. Oxford U Pr, 1985, x + 385 pp, $59. [ISBN: 0-19-503493-7] A self-contained graduate-level text covering the basic concepts of measurement and psychophysics. Concepts are discussed within the framework of measurement theory and in the language of contemporary mathematics. Students with the minimal mathematical prerequisite ("two undergraduate courses in calculus") should be prepared for some hard work. LCL

Reviewers

DFA: David F. Appleyard, Carleton; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; RD: Roger Day, St. Olaf; JD-B: John Dyer-Bennet, Carleton; SG: Steven Galovich, Carleton; BH: Bruce Hanson, St. Olaf; PH: Paul Humke, St. Olaf; KK: Kenneth Kaminsky, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; AM: Alan Magnuson, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; LN: Linda Ness, Carleton; AO: Arnold Ostebee, St. Olaf; MS: Michael Schneider, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MT: Michael Tveite, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

mands with $n = m$. Since $n \equiv m \pmod{k}$ and $n \neq m$ imply at least one of the inequalities $n \geqslant k$, $m \geqslant k$, and therefore $|G(r)| \leqslant 2B(r)$, one obtains

$$A(r) = F(r) - G(r) \leqslant F(r) + |G(r)| \leqslant M^2 + 2,$$

as required.

A few remarks are in order. The elementary nature and simplicity of this proof is purchased, of course, at the price of using the Weierstrassian power series definition of analyticity rather than the complex differentiability definition in the tradition of Cauchy-Riemann-Goursat. But that is the point: Complex numbers, convergence, and power series could be taught independently and *before* Calculus, albeit reversing present-day educational practice. Secondly, the number 2 on the right-hand side of the inequality for $A(r)$ plays no special role; it may evidently be replaced by any positive number no matter how small. Thus our proof actually yields $A(r) \leqslant M^2$; which is the best possible if one takes $M = \sup|f(z)|$. Finally, the above proof is closely related to, and is indeed the "discretized" version of, a well-known proof in which one uses averaging of $|f|^2$ with respect to arclength measure on the circle $|z| = r$. This yields directly

$$\frac{1}{2\pi}\int_0^{2\pi}|f(re^{i\theta})|^2 d\theta = A(r),$$

whence $A(r) \leqslant M^2$ is evident. The idea of replacing the integral by the finite Riemann approximating sum and thereby eliminating the last vestige of Calculus from the proof may have some value in other contexts where the needed integration theory is sufficiently cumbersome so that a more elementary procedure using sums instead of integrals has appeal.

This note is dedicated to the memory of Sándor Lénárd, uncle, engineer, and connoisseur of mathematics from whom, many years ago, the writer first heard about complex numbers.

A SELF-DUAL CARD TRICK BASED ON CONGRUENCES AND k-SHUFFLES

KURT EISEMANN
Department of Mathematics, San Diego State University, San Diego, CA 92182

Introduction. A card trick based on number-theoretic properties, resurrected by Martin Gardner [3] and analyzed in [4], has a sequel. The originator [1] has devised a second card trick [2], likewise based on congruences and k-shuffles, that exhibits the even more astonishing and extremely rare property of self-duality. The present paper describes in simplified form a generalized version of it, together with its mathematical basis. While the first trick dealt with two decks possessing mutual duality, the second is more impressive because it deals with a single deck that exhibits self-duality. It is less flexible than the first because rather than being valid for a variety of configurations, the given procedure applied to a chosen size deck furnishes essentially only a single array possessing this amazing property.

As a by-product, the notation and set-up described herein permit a further extension of the purely mathematical first card trick that significantly enhances the complexity of its aspect and its fascination.

Notation. The manner of enumerating card sequences, dealing, and the definition and execution of k-shuffles will be exactly as described in [4]. From a deck of playing cards, select cards numbered 1 (Ace) through b from each of s suits. Define the corresponding modulus as $m = bs + 1$, which we require to be prime. For illustration we shall take $b = 7$, $s = 4$, $m = 29$. Denote diamond, spade, heart and club by D_0, S_1, H_2, C_3 respectively—a redundant but mnemonic and effective notation. The 5 of hearts, for example, will be denoted $H_2 5$. This is to be interpreted as the positional notation for a 2-digit number in base b, with a *numerical value* in the decimal system of $2b + 5$ (= 19 here). The subscript of the letter designating the suit is the coefficient of b for computing the numerical value of a card. The shapes of the conventional

symbols for the suits offer a mnemonic hint to their respective subscripts. The numerical values of the chosen cards thus range from $a_1 = D_0 1 = 1$ to $a_{28} = C_3 7 = 28$ without gaps.

Deck B. Take the above deck sequenced in ascending order of numerical values ("deck A") and append a Joker (or any King) as the mth card. Have the audience cut the deck, then perform a k-shuffle with $k = 4$ exactly as described in [4], and repeat with $k = 3$ (or first $k = 3$ and then 4) "in order to mix up the cards". Cut to restore the Joker to last place and inconspicuously verify that the first card of the deck (bottom card when the deck faces upward) is $S_1 5$. Denote this sequence as "deck B". Deal deck B face down from left to right, beginning with the top row, into s $(= 4)$ rows of b $(= 7)$ cards each and discard the Joker.

Perspectives. We assume the table to separate the performer from the audience. The positions of the cards on the table are to be viewed in two different ways.

Audience perspective: Each row represents one suit, with face points 1 through b running from left to right. The suits in order, top to bottom rows as viewed by the audience, are D_0, S_1, H_2, C_3. To indicate the suit associated with each row, place to the left of each row (as viewed by the audience) the Queen of the corresponding suit face up.

Performer's perspective: Defined the same way, but as viewed by the performer; thus the audience's top left ($D_0 1$) coincides with the performer's bottom right ($C_3 7$). Explain this convention to the audience. To indicate the suit associated with each row, place to the left of each row (as viewed by the performer) the King of the corresponding suit face up.

Pointers to target. Let a spectator ask for the location of any card ("target"), e.g., "where is the 2 of clubs?". Turn the card ("pointer") in the *spectators' position* $C_3 2$ (spectators' bottom row, second column from their left) face up. Its face value $S_1 7$ gives the location of the desired target card from the *performer's perspective* (second row, rightmost column). Turn the indicated card face up and, lo!, it is the 2 of clubs! Repeat for several more cards. After the cuts and shuffles, these pointer-to-target relationships within one and the same array, subsisting for *every* card, appear utterly astounding. Indeed only one out of every 10^{22} possible arrangements possesses this property!

If you interchange the spectators' and performer's perspectives, the procedure will still be valid: In this case the card in the *performer's* position $C_3 2$ yields the pointer $H_2 1$ which, used as location in the *audience's* perspective, correctly delivers the specified target card $C_3 2$!

For a particular target card, say $C_3 4$, stress for future reference that the associated pointer (obtained from the spectators' perspective) is $H_2 5$.

Deck C. Restore deck B (with $a_1 = S_1 5$) and again append the Joker. Stating that you want "to mix up the cards more thoroughly", perform: Cuts; a 7-shuffle; cuts; a 4-shuffle; and a cut to restore the Joker to last place, resulting in "deck C". Verify that the first card of the deck (bottom card when the deck faces upward) is $H_2 3$. Deal onto the table exactly as was done with deck B. With the same conventions on performer's and audience's perspectives as before, demonstrate the reappearance of self-duality—to the audience's incredulous consternation.

For the particular target card chosen earlier for deck B, demonstrate that the associated pointer is now $S_1 3$, showing that this is a different array.

Surprise ending. Restore deck C (with $a_1 = H_2 3$) but leave off the Joker. Stating that you want "to mix up the cards still more", apply a 3-shuffle, but deal all cards onto the table face *down*, and reassemble as if the absent Joker were still at the end of the deck. Specifically, place the middle heap over the left heap and the combination on top of the right heap. Deal with faces down to 3 spectators and yourself "for a game of bridge" in which, for simplicity, "the last card dealt will be turned face up to indicate the trump suit". The result of the deal: Surprise!

Mathematical explanation. It has been shown in [4] that for a deck of m cards, with m prime,

each k-shuffle amounts to considering the deck laid out in a circle and picking cards from every kth position in sequence. For an initial deck with numerical values in natural order, shuffles with different k_j will therefore produce the sequence of face values

$$K, 2K, 3K, \ldots (\mathrm{mod}\ m), \text{ where } K = \Pi k_j.$$

It was also shown in [**4**] that a pointer-target relationship between two decks must necessarily be mutual: Whenever card P points to Q, necessarily card Q must point to P. In the present instance, however, the pointer and target decks have coalesced, so that the property of duality becomes self-duality.

What values of K will produce such self-duality? The value of the jth card of deck B, which is $a_j \equiv jK (\mathrm{mod}\ m)$, should indicate the position which, when viewed from the opposite perspective, harbors the card with value j. Hence we require that

$$a_{m-jK} = (m - jK)\, K \equiv j\ (\mathrm{mod}\ m)$$

for each j, i.e., $K^2 \equiv -1\ (\mathrm{mod}\ m)$. For $m = 29$ we therefore require $K = 12$ or 17, i.e., the first card of the shuffled deck must be either $S_1 5$ or $H_2 3$. This is precisely the result obtained respectively for decks B and C. Note that decks B and C are merely reflections of each other.

Examination of the array on the table (for deck B) shows that if we traverse it in systematic order, beginning at any corner and proceeding in the direction of the corresponding arrow sketched in Fig. 1, consecutive cards throughout the array have a constant difference in numerical values (mod 29), namely the value shown in Fig. 1 in the corner from which the traversal is started.

$S_1 5$	$12 \rightarrow$	$\downarrow 26$	$C_3 5$
$D_0 3$	$3 \uparrow$	$\leftarrow 17$	$H_2 3$

FIG. 1. Constant differences corresponding to directions of traversal. (4 suits, 7 cards each.)

Alternatives. Decks B and C may be obtained in alternative ways. First, as was shown in [**4**], the indicator property is intrinsically mutual, so that a 180° rotation of the array on the table, i.e., a reversal of the card sequence, preserves self-duality. We may therefore first aim for $K = 17 \equiv -12$ by using for k_j, for example, the factors of any member in the identities (mod 29) $17 \equiv 3 \cdot 5 \cdot 5 \equiv 8 \cdot 13 \equiv 4 \cdot 5 \cdot 11$. Similarly deck B ($K = 12$) may be reached from deck A by using the factors of any member in $12 = 3 \cdot 4 = 2 \cdot 6 \equiv 2 \cdot 5 \cdot 7 \equiv 9 \cdot 11$.

Additional k-shuffles may be interpolated by using for k_j the groups of factors in any of the identities

$$1 \equiv 5 \cdot 6 = 3 \cdot 10 \equiv 8 \cdot 11 \equiv 9 \cdot 13 \equiv 3 \cdot 5 \cdot 5 \cdot 12$$

or

$$-1 \equiv 4 \cdot 7 \equiv 4 \cdot 6 \cdot 6 \equiv 3 \cdot 7 \cdot 11 \equiv 5 \cdot 9 \cdot 9 \equiv 5 \cdot 10 \cdot 11.$$

Other variations consist in reassembling the array on the table *columnwise*, the value of the first card being identical to the common difference shown in Fig. 1, viz., $a_1 = D_0 3 = 3$ or $C_3 5 = 26 \equiv -3$. One may then recreate the array (which requires $K = \pm 3 \cdot 4$) by applying k-shuffles with factors drawn from

$$4 \equiv 3 \cdot 11 \equiv 7 \cdot 13 \equiv 4 \cdot 1$$

or from

$$-4 \equiv 5 \cdot 5 \equiv 3 \cdot 3 \cdot 6 \equiv 2 \cdot 7 \cdot 8 \equiv (-4) \cdot 1 = 4 \cdot (-1),$$

where 1 or -1 are replaced by factors shown in the preceding paragraph. Alternatively one may reassemble columnwise and choose k so as to lead directly into the ending.

As yet another variation, one may shuffle to yield

$$a_1 = 3 \equiv 4 \cdot 8 \equiv 3 \cdot 5 \cdot 6 \equiv 3 \cdot 8 \cdot 11 \equiv 3 \cdot 5 \cdot 5 \cdot 7$$

or

$$a_1 = 26 \equiv 5 \cdot 11 \equiv 3 \cdot 4 \cdot 7 \equiv 2 \cdot 4 \cdot 5 \cdot 5 \equiv 7 \cdot 7 \cdot 10 \equiv 7 \cdot 9 \cdot 11$$

and deal onto the table columnwise (7 columns of 4 cards each) in either of two ways: downward (beginning with the rightmost column), or upward (beginning with the leftmost column), as implied by the arrows in Fig. 1.

Generalization. The above can be generalized to any values of the parameters b and s, provided the resulting modulus $m = bs + 1$ is prime. For a full deck of playing cards ($b = 13$, $s = 4$, $m = 53$) including the face values X(for 10), J, Q, K, dealings become much more tedious, the positional notation (such as $H_2K = 39$) a little more esoteric, and the list of factors harder to memorize. For the self-dual table we need as before $a_1^2 \equiv -1 \pmod{53}$, leading here to $a_1 = S_1X(= 23)$ or $H_24(= 30)$. The constant differences (mod 53) for traversal of the corresponding array from any corner are shown in Fig. 2, while the factors needed by the performer are $30 = 5 \cdot 6$ and $23 \equiv 4 \cdot 8 \cdot 9$ or other congruential equivalents mod 53.

S_1X	23 →	↓34	H_28
S_16	19↑	← 30	H_24

FIG. 2. Constant differences corresponding to directions of traversal. (Full deck.)

The reader should have no difficulty now explaining the mathematical reason for the outcome of the surprise ending; or working out (and testing) the k-values required to achieve the same ending with a full deck of cards; or deriving k-values for the indicated variations for a full deck of cards or for any other parameters b, s with prime m.

Further extension of the first card trick. The notation above for the numerical values of cards and for positions in a 2-dimensional array can be applied to make the first card trick of [**4**] even more stunning: Rather than dealing with only a single suit for each of the two decks, use 4 suits with face values 1 through $b(= 7$, say) for each deck. The terms "black deck" or, equivalently, "blue deck" (the successive power residues) and "red deck" (the corresponding pointers) to distinguish the two decks will now refer not to the suits on the face of the cards but to the color on their backside.

Begin with a blue deck sequenced in successive powers modulo 29 of its primitive root 2, interpreted in the notation of the present paper (see Fig. 3, in which the subscripts have been omitted), and a corresponding red deck of pointers (Fig. 3) to which a Joker (or any King) has been appended.

Rows	The black (or blue) deck							The red deck							
Diamonds	$D2$	$D4$	$S1$	$H2$	$D3$	$D6$	$S5$	$C7$	$D1$	$D5$	$D2$	$C1$	$D6$	$S5$	
Spades	$C3$	$H5$	$S2$	$H4$	$D7$	$S7$	$C7$	$D3$	$S3$	$C2$	$C4$	$D7$	$H4$	$S6$	
Hearts	$C6$	$C4$	$H7$	$S6$	$C5$	$C2$	$H3$	$C6$	$D4$	$H7$	$S4$	$S2$	$C3$	$H3$	
Clubs	$D5$	$S3$	$H6$	$S4$	$C1$	$H1$	$D1$	$C5$	$H6$	$S1$	$H2$	$H5$	$H1$	$S7$	J

FIG. 3. Initial sequences of pair of decks for the extended first card trick.

Perform the cuts and k-shuffles of the first trick any number of times with *arbitrary* k. However, in order to avoid errors in reassembly, it is expedient to limit each k to the range $1 < k < 8$. Apply the requisite final cut to the red deck, and in it determine the position number of the Ace of diamonds ($D_01 = 1$). Convert its position number into our 2-place base b notation and cut the *blue* deck so as to bring the corresponding card to first place. Deal the blue deck row-wise face down into a 4×7 array, then deal the red deck face down in the same manner on

top of the same array. Here the table is viewed from but a single perspective both by the audience and the performer. For illustration we shall refer to the initial decks of Fig. 3.

Let a spectator name any card (1 to 7) of any suit; say it is $C4$ (the 4 of clubs). In the named position (bottom row, 4th card from left), the *red* card ($H2$ in Fig. 3) will give the location of $C4$ of the *blue* deck, while the *blue* card ($S4$ in Fig. 3) will give the location of $C4$ of the *red* deck!

As described for the first trick, restoration of the two decks and continued k-shuffles preserves reciprocity, possessed by no fewer than $(m-1)\phi(m-1)=336$ different arrangements of the above two decks (1248 different arrangements for a pair of full decks with $b=13$). Now, however, each deck contains cards from *all* 4 *suits* and is displayed in the form of a 2-dimensional array, with positions identified through the positional notation described for the second trick. We have thereby introduced an additional dimension, viz., the suits, to each of the decks—both for the faces of the cards and for their *positions*. The connection between suits and the system of numbers in base b remains entirely veiled from view, heightening the mystery of the underlying number-theoretic basis. In this more general form, the highly mathematical first trick is therefore apt to come across as even more complicated and astonishing than in its original form.

References

1. Charles Sanders Peirce, Collected Papers, Harvard Univ. Press, 1933, vol. 4, p. 473, The First Curiosity.
2. Ibid, p. 543, A Second Curiosity.
3. Martin Gardner, Mathematical Games, Scientific American, July 1978.
4. Kurt Eisemann, Number-theoretic analysis and extensions of "The most complicated and fantastic card trick ever invented", this MONTHLY 91 (1984) 284.

THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

For instructions about submitting material for publication in this department see the inside front cover.

ON CATERPILLARS, TREES, AND STOCHASTIC PROCESSES
An essay on learning to be mathematically creative

JOHN C. TURNER
Department of Mathematics, University of Waikato, Hamilton, New Zealand

Prologue. This paper is addressed, in the main, to high school and undergraduate mathematics students and high school teachers. We are sure that most students have no idea that new mathematics is constantly being invented, and that most fields of mathematics are expanding at a great rate. Still fewer students (and few high school teachers too) would believe that they could invent new mathematics; they would not know how to set about the task.

We hope that the following discussion on how to amalgamate a variety of objects and mathematical ideas will inspire some of them to try their hands at problem inventing and solving. In short, to become mathematically creative. We assure them that studying one's own problems is many-fold more pleasurable than tackling those proposed in textbooks. There are few joys to match that experienced when an attractive, unsuspected relationship is discovered, or when a proof to confirm a suspected theorem is finally nailed down. One's own theorem!

We warn, however, that the process is messy; work has to be intense and protracted; and answers to self-generated problems may not even be there to be found!

1. Introduction. In textbooks on probability theory, elementary problems that are discussed

usually involve coin-spinning, dice-shaking experiments, card shuffling and dealing, gambling games, or random selections of coloured balls from urns (Grecian or otherwise). In this paper we put together a much more congenial mix of objects, about which a virtually unending list of interesting probability problems can be posed. Not all, however, will admit of easy solution. Nevertheless, a student who attempts to solve them, and make up further similar problems for himself (or herself), will find it easy and pleasant to keep the objects in mind, and at the same time will find many interesting relationships between such things as Fibonacci numbers, binomial and other probability laws, transition matrices of Markov chains, recurrence relations, and so on.

As the title of our paper suggests, our problems will relate to caterpillars and trees. Later we shall indicate how we may broaden our scope (and fantasies) by introducing other fauna and flora, such as spiders, snakes and rambling vines.

Everyone knows what a *caterpillar* is, and what a *tree* is. Some students will know what a mathematician's definition of a tree is, namely a connected graph with no cycles. Almost no one knows what a *stochastic process* is! Let us begin, then, with a descriptive definition of such an entity:

A *stochastic process* is a mathematical model of some natural process whose various components are governed by probability laws.

The word *stochastic* is often used synonymously with *probabilistic*, or with *random*, and usually a stochastic process is a model of something that is happening over a period of time. If time is involved, we say that the stochastic process is time-dependent.

Two simple examples of natural processes which can be described by stochastic processes are: the number of persons queueing at a supermarket cash desk (queue theory); the output of a message source which produces one of several types of message (communications theory).

Let us now take a simple problem, and see how it leads us into stochastic processes and various other realms of mathematics.

PROBLEM. A bird carrying a caterpillar in its beak flies over a tree. The caterpillar jerks free, and falls into the tree onto a node (a vertex of the tree).

What is the expected number of leaves the caterpillar finds above her (our caterpillar is a female) in the portion of the tree stemming from her landing point?

It is easy to see that the problem is ill-posed. A lot more things have to be said before anyone can take it seriously, and get down to the business of seeking a solution. (As an aside, this is often the situation when a scientist is doing research into some natural process. His first questions about the process are usually ill-posed. Before he can begin formulating mathematical models he has to refine the questions, and make hypotheses or assumptions about all manner of details.)

2. Fixing the problem (i.e., making it well-posed). In order to describe what has happened in mathematical terms, we must specify:

(i) the kind of tree the caterpillar falls into, and
(ii) the law by which its landing point is selected.

We will settle these requirements by choosing a nice kind of tree to place under the caterpillar, namely a Fibonacci tree, and by stating that she falls in a completely random fashion onto the set of nodes of the tree (i.e., if there are a total of N nodes, any particular node is landed upon with probability $1/N$).

So what are Fibonacci trees? This is best answered by showing several. They are defined in sequence, and the diagrams in Fig. 1 are of the first six.

Notice that successive numbers of leaf-nodes in the trees, after T_1 and T_2, are 2, 3, 5 and 8, which are Fibonacci numbers. Note too that each tree after the second is constructed by mounting the two previous trees on a fork, as shown in Fig. 2.

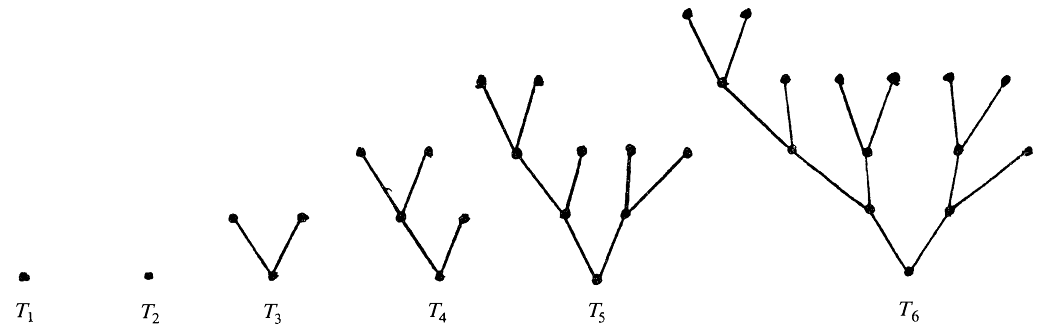


FIG. 1. The first six Fibonacci trees.

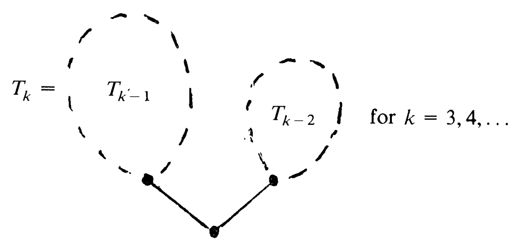


FIG. 2

It is clear that this will serve as a recursive definition for an infinite sequence of trees $T_1, T_2, \ldots, T_k, \ldots$. Then the tree T_k with $k > 2$ will have F_k lead-nodes, since the Fibonacci number sequence is defined in the same recursive way, namely $F_k = F_{k-1} + F_{k-2}$ with starting values $F_1 = 1$ and $F_2 = 1$.

We shall need to know several properties of Fibonacci numbers to solve our stochastic problem. The student might like to prove each property we use, or find its derivation in [**2**] or any elementary text on number theory. We shall also need formulae relating to the trees $\{T_k\}$. First let us find a formula for the total number N_k of nodes in the kth Fibonacci tree. We shall exploit the recursive definition of T_k to find it; the technique illustrates a method that can often be used with counting problems defined on tree sequences such as ours.

PROPOSITION 1. *The total number N_k of nodes in tree T_k is $2F_k - 1$.*

Proof. By the recursive definition of T_k, we see that

$$N_k = 1 + N_{k-1} + N_{k-2}.$$

The 1 on the RHS refers to the root node. Adding 1 to both sides, we can write

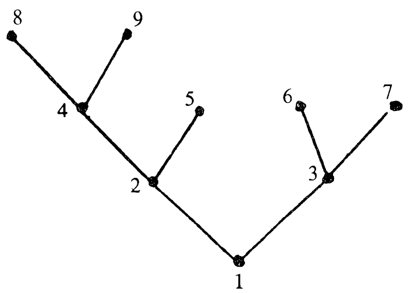
$$(N_k + 1) = (N_{k-1} + 1) + (N_{k-2} + 1).$$

Letting $M_k = N_k + 1$, and noting that $M_1 = M_2 = 2$, we see that the sequence $\{M_k\}$ is $\{2, 2, 4, 6, 10, \ldots, 2F_k, \ldots\}$. Therefore

$$N_k = M_k - 1 = 2F_k - 1, \quad \text{with } N_1 = N_2 = 1. \tag{1}$$

3. Solving the caterpillar problem. We shall tackle our problem in two stages. First we will take an easy particular case, letting the caterpillar drop on a small tree: we will use T_5 for this. Then we will attempt to find a general solution using T_k. This procedure, by the way, is the way that much new mathematics is found. Scrabble away at simple cases, to obtain familiarity with the concepts involved; then gradually forge methods to deal with generalisations.

The picture below shows T_5 with its nodes labelled $1, \ldots, 9$, and the table alongside tells much of the solution story.



T_5

Node landed upon	Number of leaves above that node
1	5
2	3
3	2
4	2
5	0
6	0
7	0
8	0
9	0

Letting $L \equiv$ number of leaves above the node, and collecting the information into a probability distribution table, we obtain the following:

L	0	2	3	5
$p(L)$	$\frac{5}{9}$	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{1}{9}$

To get each $p(L)$ value we have used the fact that each node has chance $1/9$ of being landed upon. Then, for example, the chance of landing on a leaf-node (i.e., $L = 0$) is equal to the chance of landing on one of nodes 5, 6, 7, 8, or 9, which is $5/9$; and similarly for the other L-values.

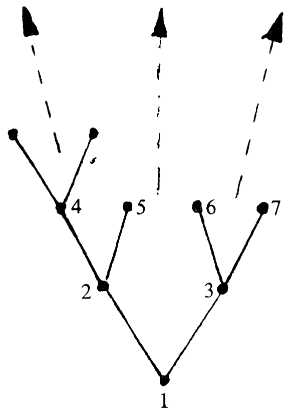
Now, at last, we can answer our problem. For tree T_5 we have:

$$E[L] = \sum Lp(L) = 0 \cdot \frac{5}{9} + 2 \cdot \frac{2}{9} + 3 \cdot \frac{1}{9} + 5 \cdot \frac{1}{9} = \frac{12}{9}. \tag{2}$$

Thus the caterpillar may expect to have $1\frac{1}{3}$ leaf nodes above her when she lands on the tree.

The particular case was easy. Can we find a general solution, for T_k?

Note first (and take pleasure!) that all the digits occurring in the probability distribution table are Fibonacci numbers except the denominator 9. Note also, from the tree diagrams and the recursive definition of the trees, that each node of a Fibonacci tree is the root of a smaller Fibonacci tree growing above it. So the general case, in tabular form, becomes:



T_k

Node landed upon (x), and the tree it supports	Number of leaves above x L
1 T_k	F_k
2 T_{k-1}	F_{k-1}
3 T_{k-2}	F_{k-2}
4 T_{k-2}	F_{k-2}
5 T_{k-3}	F_{k-3}
6 T_{k-3}	F_{k-3}
7 T_{k-4}	F_{k-4}
$\vdots$	$\vdots$
N_k T_1	0

It is easy to show the following (the student is invited to prove each statement):

The number of occurrences of $L = 0$ is F_k.

The number of occurrences of $L = F_r$ is F_{k-r+1}, for $r = 3, \ldots, k$.

Armed with this information the general case* is seen to have solution

$$\text{(3)} \qquad \text{mean } \mu_L = E_k[L] = \frac{2F_{k-2} + 3F_{k-3} + 5F_{k-4} + \cdots + F_k \cdot 1}{N_k} = \frac{\sum_{r=3}^{k} F_r F_{k-r+1}}{2F_k - 1}, \quad \text{using (1) for } N_k.$$

If we introduce the symbol s_k for the sum in the numerator of (3), we can write this as

$$\text{(4)} \qquad E_k[L] = \frac{s_k}{2F_k - 1}.$$

The student may care to study the sum s_k. It may be possible to find some neat closed form for it, using identities for Fibonacci numbers. It is easy to show that it satisfies the recurrence equation

$$\text{(5)} \qquad s_k = s_{k-1} + s_{k-2} + F_k, \quad k \geqslant 2.$$

Using this we can show the polynomial generating function for s_k to be

$$\text{(6)} \qquad G(x) \equiv \sum_{k=0}^{\infty} s_k x^k = \frac{x^2(2 + x)}{(1 - x - x^2)^2}.$$

Expanding the RHS as a polynomial in x, and checking that the first few coefficients of x^k are in fact s_k values, would be a useful exercise.

Equation (11) below gives another useful result for s_k.

Before leaving our study of μ_L, let us ask the student to consider whether an alternative method of solution can be found. A creative mathematician might reason thus: "Suppose I find another way of getting μ_L. That will give me a new method to work with; and it might even give me a different formula for μ_L. If the latter, I can equate it to (4), and hence obtain an identity for s_k. And *that* might be a new result in Fibonacci numbers." And so his mind would run on, seeking new ways of studying the situation and all its ramifications.

4. Further properties of the distribution of L. We have found an expression for the mean of the distribution of L. What about some other parameters of the distribution? The median, say; or the variance?

One easily finds (and experiences a certain measure of shock in doing so) that the *median* of L is always zero, no matter what value of k is chosen! The median is that value of L which has half the probability distribution to its left and half to its right. The property just claimed for the median follows immediately from the fact that

$$\text{(7)} \qquad P[L = 0] = \frac{F_k}{2F_k - 1} = \frac{1}{2 - \frac{1}{F_k}} > 1/2.$$

---

*We should now introduce a more careful notation, to indicate that the distribution of L is conditional upon the tree T_k upon which the caterpillar falls. For example, we could use for the mean $\mu_L(T_k) \equiv E[L|T_k]$. For simplicity, however, we shall use only μ_L or $E_k[L]$.

That is, no matter what size of tree is taken, more than half the total probability is concentrated at the origin. The shock is soothed when one realises that all this means is that more than half of the nodes of any Fibonacci tree are leaf-nodes.

This leads us to think of the relationship of the mean to the median, for *mean minus median* is a measure of skewness. Clearly this is always positive for the L-distribution. It is usual to standardize this measure, dividing it by the standard deviation of the distribution. As the median is zero, this gives us the parameter μ_L/σ_L to investigate.

Using the table given above (2), and the results below it, we can write down an expression for the variance, thus:

$$\sigma_L^2 \equiv V_k[L] \equiv E\left[(L-\mu_L)^2\right] = \frac{1}{2F_k-1}\sum_{r=3}^{k} F_{k-r+1}(F_r-\mu_L)^2. \tag{8}$$

The student may care to seek simple (i.e., nice) formulae for μ_L/σ_L and σ_L^2, using (3). We don't know whether any exist.

5. Asymptotic results. A natural question to ask (we have already asked and answered it for the median, see (7)) about mathematical results relating to a system is: What happens when a system parameter tends to infinity? In our problem we are thinking of letting $k \to \infty$; that is, of what happens to the distribution of L and its parameters when the Fibonacci tree in the model is allowed to grow without limit.

The behaviour or form of a parameter 'at' or 'on its way to' infinity is often much simpler than its finite behaviour or form. We gain insight into processes by studying these *limit forms*: we may call information gleaned from them *asymptotic results*.

We will show that μ_L tends to infinity in a nice way, with k. In fact, the curve of μ_L, plotted against $k = 3, 4, \ldots,$ comes closer and closer to a straight line, which we call the *asymptote* of the curve.

PROPOSITION 2. *The asymptote of $\mu_L = E_k[L]$ is the line $y = mk + c$, with the gradient and the intercept respectively being*

$$m = \frac{\alpha}{2\sqrt{5}}, \quad \text{and} \quad c = \frac{1}{10}(3\beta - 2\sqrt{5}\,\alpha + 3), \tag{9}$$

$$\text{where } \alpha = \frac{1+\sqrt{5}}{2} \quad \text{and} \quad \beta = \frac{1-\sqrt{5}}{2}.$$

Proof. First we remark that α and β given in (9) are the roots of $x^2 - x - 1 = 0$ and that the kth Fibonacci number may be expressed thus:

$$F_k = \frac{1}{\sqrt{5}}(\alpha^k - \beta^k). \tag{10}$$

(Prove this by inserting it in the recurrence equation for F_k; or see [**2**].)

Using (10) we can show with a little algebra (we give only a few of the steps) that:

$$\begin{aligned} s_k &= \sum_{r=3}^{k} F_r F_{k-r+1} = \sum_{i=0}^{k-3} F_{i+3}F_{k-i-2} \\ &= \frac{1}{5}(k-2)(\alpha^{k+1}+\beta^{k+1}) - \frac{1}{5}\left[\beta^3\alpha^{k-2}\sum_{i=0}^{k-3}\left(\frac{\beta}{\alpha}\right)^i + \alpha^3\beta^{k-2}\sum_{i=0}^{k-3}\left(\frac{\alpha}{\beta}\right)^i\right] \\ &= \frac{1}{5}(k-2)(\alpha^{k+1}+\beta^{k+1}) - \frac{1}{5}\alpha\beta(\alpha^2+\beta^2)\frac{(\alpha^{k-2}-\beta^{k-2})}{(\alpha-\beta)}. \end{aligned} \tag{11}$$

If we now divide through by $2F_k - 1$, and use (10) to express F_k in terms of α and β, we get μ_L as a function of α and β.

Dividing both numerator and denominator by α^k, we get an expression from which the behaviour of the mean as $n \to \infty$ can easily be deduced, using the fact that $\lim_{k\to\infty} \left(\frac{\beta}{\alpha}\right)^k = 0$, and $\left(\frac{\beta}{\alpha}\right)^k$ tends to zero faster than k tends to infinity. If the student will fill in the details, he will see that Proposition 2 has been proved.

Another way to study the gradient of μ_L (that is, its rate of increase with respect to k), is to study the difference

$$\Delta_k = \frac{s_{k+1}}{F_{k+1}} - \frac{s_k}{F_k}. \tag{12}$$

Using properties of the Fibonacci numbers, and some algebra, we can show that:

$$\lim_{k\to\infty} (\text{gradient of } \mu_L) = \frac{1}{2} \lim_{k\to\infty} \Delta_k = \frac{\alpha}{2\sqrt{5}}, \tag{13}$$

as in (9).

Consider also using (5), thus: Divide through by F_k and take limits of both sides. Does it help?

6. Further developments and new problems. The student is now invited to join in the fun—and invent new mathematics. He/she must spend time thinking about the situation, and deciding how it can be changed to offer new possibilities for study. Let ideas for various 'caterpillar adventures' flow freely, and write down questions about them. Each adventure will suggest a possible mathematical model; efforts to define that model, to symbolise it and solve the resulting combinatoric problems or equations, will lead to interesting relationships, discovery of algorithms, and so on. There will be endless possibilities for using micro-computers to study the various series, parameters and functions, when efforts to find neat formulae founder.

Doing creative mathematics is a hunting activity. It is one that requires boundless curiosity and tenacity; the search for answers, and methods to acquire answers, must go on endlessly, once problems concerning an envisaged situation have been specified.

This kind of activity is a long cry from filling in formulae and following routine algorithms drilled into students by teachers. It is an activity that is rarely encouraged in the classroom; we hope that this paper will encourage a few students and teachers to attempt it.

Let us write down some ideas and problems that spring easily to mind, which could set a student along the royal road to creativity:

(1) Consider the caterpillar's climbing activities, as she proceeds up to a leaf. How many branches does she cover before reaching the leaf? Which leaf? What branching rules is she to use? (Left every time? Or right? Or left with probability p? Or ...?).

(2) What is the path length distribution, given the node landed upon? How does it depend on the rules 'fixed' for branching? What is the expected path length to a leaf? Can you make twig lengths different; can you specify them in some way that allows you to find interesting solutions (or any solutions at all!) to your problems?

(3) Can you change the tree structure? There are many ways of defining number sequences; look them up, and invent a tree-sequence evolution algebra, incorporating the number sequences in some way. You might find this easier than you think—you will enjoy inventing an algebra. Then solve again, if you can, all your probability problems for the falling caterpillar, on the new tree structure.

(4) Do you have to stay with tree structures? The caterpillar might land on a wire-netting fence, for example!

(5) Why study only *one* caterpillar? Let the bird drop two caterpillars. Then the number of problems available is surely the square or cube of the number for one caterpillar. For example, where do they land? What is their expected distance apart? How does this distance change as they begin to crawl up (why not down... investigate the 'up-down' duality possibilities in all your problems) according to the rules you lay down for them? Do your problems get (a) harder, (b) more interesting, with *three* caterpillars? With n caterpillars?

(6) When you are tired of just caterpillars, bring in a few distractions and obstacles. A strategically placed spider, capable of spinning a web in such-and-such a way (your decisions) would create interesting problems for the caterpillars, and headaches for you.

And how about including jumping frogs, preying bats, insecticides, rotten twigs, caterpillar-eating flowers, woodmen with axes...?

Go to it! Let your imagination roam. But beware: too many snags, too many rules, and you will find that your mathematical models will become too complicated for you to do anything with. That, however, will provide a good lesson for you to learn: that mathematical methods are severely limited. Compromises have to be made between the desire for complexity of models, and the necessity of mathematical tractability if analytic solutions are to be achieved. If you find such compromises too frustrating, you will have to learn about simulation methods, and begin computer-aided modelling. Then you will become a computer scientist, and beyond the scope of this paper!

7. Relevance. We can imagine that many mathematicians (if they were to read this paper) would say of our proposals: Why not have your students invent some problems about real world situations? Surely there are plenty of them? They would be thinking no doubt of problems of fluid flow and traffic jams, quality control, meteorology, packet switching, information processing, etc., etc.

Our reply would begin with a statement that all mathematical models are abstractions—dream worlds. Further, our guess is that few mathematics papers, whether 'pure' or 'applied', are *truly* applicable to a real world situation. We wonder how many mathematicians study mathematical problems because they want to see the solutions applied, except perhaps to further mathematical problems! To paraphrase many of the great mathematicians of the past on this matter, the joys of creating mathematics lie in the excitement of the hunt and the beauty of the catch:

Where there is mathematics, there is beauty. PROCLUS

Thinking about caterpillars on trees is, after all, just a device to capture a student's imagination. Once captured, the student will create his own problems, and, seeking to solve them, he or she will be led quite gently into several fields of mathematical inquiry. These will include number theory, graphs, combinatorics, probability and Markov chains, and recurrence equations. For a student to link together ideas drawn from each of these, under his own volition, will be no mean accomplishment, stemming from such a simple device.

It has occurred to us that the dream world he will be imagining is very like the present-day computer games that hold so many people in total absorption. Though this paper was not inspired by thoughts of computer games, it would seem fitting to end by dedicating it to them. For we believe that much mathematics of the future (and methods of studying and applying it) lie with a melding of symbolic language processing and computer modelling. Students reared on caterpillar-tree problems might just have a head start for this future!

For those wishing to know whether studies of stochastic processes on trees have been carried out by others, and whether they might have some useful applications, the answer is: Yes, they have. There have been many applications already in the fields of computer data storage and information processing [**1**]. Reference [**3**] contains two papers on stochastic generation of tree-structures; and [**4**] is a very recent paper dealing with random motion on binary trees, with applications cited on movement of nutrients in root systems of plants.

8. The Caterpillar Club. To encourage creative mathematical activity along the lines suggested above, we hereby inaugurate *The Caterpillar Club*. To become a member you must submit to us (typed or written up neatly) a caterpillar adventure model of your own devising, with some interesting mathematical problem or problems worked out in relation to it. If it is satisfactory, we will send you an illuminated membership scroll: and from time to time we will send out a Newsletter to members, listing problems and results achieved.

If a sufficient number of contributions is forthcoming, we will attempt a classification, with indications of 'real' applications; the resulting document could prove useful to teachers and students, perhaps even to researchers.

Epilogue. If a teacher is to use the ideas of this paper, he must consider how to put them before his students. The problem-inventing and -solving projects could be placed alongside a course in discrete mathematics (if it included some elementary probability) but it would need very careful organisation, and all the necessary reference materials would have to be on hand.

We feel bound to conclude on a pessimistic note. We know that placing ideas and materials in front of students is one thing; inspiring them to reach the flash point where intense commitment to solving a problem sets in is quite another.

Maybe it is unreasonable to expect any student, labouring on the grade-point-average treadmill, to be able to supply such a commitment. And maybe *that* says something about our educational systems!

References

1. Y. Horibe, Notes on Fibonacci trees and their optimality, Fibonacci Quarterly, May 1983, 118–128.
2. N. N. Vorob'ev, Fibonacci Numbers, Pergamon Press, 1961.
3. E. F. Harding and D. G. Kendall (Edrs.), Stochastic Geometry, Wiley, 1974.
4. D. H. Frank and S. Durham, Random Motion on Binary Trees, J. Appl. Prob., 21 (1984) 58–69.

WHY SQUARE ROOTS ARE IRRATIONAL*

WILLIAM C. WATERHOUSE

Department of Mathematics, Pennsylvania State University, University Park, PA 16802

Like many lecturers, I find it useful to begin a discussion of real numbers and approximation by making the students realize that many "familiar" numbers like $\sqrt{2}$ are not expressible by fractions. To convince them of this fact, I have to give them some explanation, as they are not likely to have considered the question before. I can choose any one of several proofs of the irrationality of $\sqrt{2}$. But in practice none of them really fits well into a situation like this. The problem is that every one of them is rather too "tricky", too indirect; even if the students manage to follow such a proof, it distracts them from the major point of the lecture. What the situation requires is an argument that will make the fact seem "obvious" to the students without requiring too much thought from them. In this note I want to explain an approach that seems to solve this problem fairly well.

We should first analyze why beginning students find the proof tricky. For one thing, irrationality by its very nature is a negative statement, and a proof of irrationality is likely to depend on some sort of argument by contradiction. In addition to that, the proof is likely to involve reasoning about unique factorization (at least for the factor 2); and though students have done explicit factorizations, they are liable to be uneasy with abstract arguments about possible and impossible factors. The approach I will present here is designed specifically to minimize these two difficulties. Of course the argument is familiar in essence (cf. [**3**, p. 34]) but this way of presenting it will allow us to skim over the tricky part.

*This work was supported in part by the National Science Foundation, Grant DMS 8400649.

PROBLEMS AND SOLUTIONS

EDITED BY G. L. ALEXANDERSON, H. M. W. EDGAR (ELEMENTARY PROBLEMS), DALE H. MUGLER, AND KENNETH B. STOLARSKY (ADVANCED PROBLEMS)

For instructions about submitting ***proposed*** *problems for publication in this department see the inside front cover. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

Solutions *should be sent to the address given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

For instructions about submitting solutions of these Elementary Problems, which should be mailed by July 31, 1986, see the inside front cover. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgement).

E 3135. *Proposed by Hüseyin Demir, Middle East Technical University, Ankara, Turkey.*

For a scalene triangle ABC inscribed in a circle, prove that there is a point D on the arc of the circle opposite to some vertex whose distance from this vertex is the sum of its distances from the other two vertices.

Show how D may be constructed with straightedge and compass.

E 3136. *Proposed by D. M. Bloom, Brooklyn College of CUNY.*

The 1982 Putnam Exam proposed the following problem: If a, b, c, d are positive integers with $a + c = 1982$, and if $r = 1 - a/b - c/d > 0$, show that $r > (1983)^{-3}$. The stated lower bound for r is not best possible.

(a) Find the actual minimum value of r.

(b) More generally, if $N > 1$ is a positive integer, find, as a function of N, the minimum positive value of $r = 1 - a/b - c/d$ over all positive integers a, b, c, d such that $a + c = N$.

E 3137. *Proposed by Solomon W. Golomb, University of Southern California.*

It is well known that 30 is the largest integer N such that all k with $1 < k < N$ and $(k, N) = 1$ are prime. What is the largest odd integer N such that every odd integer k with $1 < k < N$ and $(k, N) = 1$ is a prime?

E 3138. *Proposed by W. R. Utz, University of Missouri, Columbia, MO.*

Let $T(n)$ denote the nth triangular number, i.e., $T(n) = n(n+1)/2$.

(a) Show that there exist infinite sequences a_n and b_n such that

$$T(2)T(a_n) = T(b_n).$$

(b) Find pairs of triangular numbers whose product is triangular such that each member of the

pair is arbitrarily large.

(c)* What is the asymptotic density of "composite" triangular numbers among all triangular numbers?

E 3139. *Proposed by Andrew Vince, University of Florida.*

Let $G_1, \dots, G_N$ be closed plane convex sets, each of diameter less than or equal to 1 and such that $\bigcup_{i=m}^{n} G_i$ is convex for all (m, n) with $1 \leqslant m \leqslant n \leqslant N$. Prove or disprove: the union of the G_i can contain no disk of diameter greater than $\sqrt{3}$.

E 3140. *Proposed by Khristo Boyadzhiev, University of Sofia, Bulgaria.*

Let $K = (\ln\sqrt{\sqrt{2} + 1})^2$.

(a) Show $\displaystyle\int_1^\infty \frac{\ln(x + \sqrt{x^2 - 1})}{x(1 + x^2)}\,dx = K.$

(b) Also show that (i) $\displaystyle\int_1^\infty \frac{\arctan x}{x\sqrt{x^2 - 1}}\,dx = \pi^2/8 + K,$

(ii) $\displaystyle\int_0^1 \frac{\arctan x}{\sqrt{1 - x^2}}\,dx = \pi^2/8 - K.$

SOLUTIONS OF ELEMENTARY PROBLEMS

Two Closely Related Modular Counting Problems

E 2943 [1982, 274]. *Proposed by Eric Anderson, Student, St. Olaf College.*

For $n \geqslant 3$, let $Z_n = \{0, 1, 2, \dots, n - 1\}$,

$$A_n = \{(a, b, c) \mid a, b, c \in Z_n, a < b < c, a + b + c \equiv 0 \pmod{n}\},$$
$$B_n = \{(a, b, c) \mid a, b, c \in Z_n, a \leqslant b \leqslant c, a + b + c \equiv 0 \pmod{n}\}.$$

Let a_n and b_n be the number of elements of A_n and B_n respectively.

a. Show that $a_{n+3} = b_n$ and $b_n = a_n + n$.
b. Find an explicit formula for a_n in terms of n.

Solution by J. C. Binz, Universität Bern, Switzerland. The map $f\colon A_{n+3} \to B_n$ $(n \geqslant 3)$ with

$$f(a, b, c) = \begin{cases} (0,0,0) & \text{if } a = 0,\ b = 1,\ c = n + 2, \\ (a, b - 1, c - 2) & \text{if } a + b + c = n + 3,\ c \leqslant n + 1, \\ (a - 1, b - 2, c - 3) & \text{if } a + b + c = 2n + 6, \end{cases}$$

is a bijection, and thus $a_{n+3} = b_n$. In B_n are all the triples of A_n, and in addition $(0,0,0)$, the $\left[\frac{n}{2}\right]$ triples containing $k, k, n - 2k$ $(1 \leqslant k \leqslant \frac{n}{2})$, and $\left[\frac{n-1}{2}\right]$ triples containing $k, k, 2n - 2k$ $(n - 1 \geqslant k > \frac{n}{2})$. Hence

$$b_n = a_n + 1 + \left[\frac{n}{2}\right] + \left[\frac{n-1}{2}\right] = a_n + n.$$

By induction (or with formal power series) one proves, starting with $a_{n+3} = a_n + n$, the formulas

$$a_n = \begin{cases} \dfrac{1}{3}\dbinom{n-1}{2} & \text{if } n \equiv 1,2 \pmod 3, \\ 1 + \dfrac{n(n-3)}{6} & \text{if } n \equiv 0 \pmod 3, \end{cases}$$

or shorter $a_n = \left[\dfrac{n^2 - 3n + 6}{6}\right]$.

Several solvers noted that the hypothesis $n \geqslant 3$ may be relaxed to $n \geqslant 1$. M. Vowe referred to P. Bachmann's *Niedere Zahlentheorie*, Vol. II, Chelsea, New York 1968.

Also solved by 27 other readers and the proposer.

Sign Relations Between Left and Right Derivatives

E 2964 [1982, 594]. *Proposed by M. Slater, University of Bristol, U.K.*

Suppose f is a real function having a left-hand derivative f'_L and a right-hand derivative f'_R at every point of the real interval I. We say that real numbers a, b *have opposite sign* provided $ab \leqslant 0$. Show that the following are equivalent:

(A) *One* of the functions f'_L, f'_R takes values of opposite sign at some *two* points p, q of I.
(B) The *two* functions f'_L, f'_R take values of opposite sign at some *one* point r of I.

Solution by Roy O. Davies, The University, Leicester, England. It is understood that f'_L and f'_R are finite everywhere, so f is continuous. If f is increasing, then both f'_L and f'_R are non-negative: (A) and (B) are both true if one is 0 somewhere, and both false otherwise; similarly if f is decreasing. If f is neither increasing nor decreasing, then its graph has chords of both strictly positive slope and strictly negative slope; therefore each of f'_L and f'_R takes values greater than zero and values less than zero, and thus (A) holds, by theorems on pp. 122–124 of R. P. Boas, *A Primer of Real Functions*, MAA, 1960; while f admits a local extremum somewhere and thus (B) holds.

Also solved by E. S. Landau, V. D. Mascioni (Switzerland), W. A. Newcomb, Chin-chi Shan, A. Villani (Italy), Pei Yuan Wu (Taiwan), and the proposer.

Another Appearance of the Catalan Numbers

E 2972 [1982, 698]. *Proposed by J. O. Shallit, University of California, Berkeley.*

Let $p(x_1, x_2, \ldots, x_n)$ be a polynomial in n variables with constant term 0, and let $\#(p)$ denote the number of distinct terms in p after terms with like exponents have been collected. Thus for example $\#((x_1 + x_2)^5) = 6$.

Find a formula for $\#(q_n)$ where

$$q_n = x_1(x_1 + x_2)(x_1 + x_2 + x_3) \cdots (x_1 + \cdots + x_n).$$

Solution by David M. Wells, Pennsylvania State University, New Kensington. We will show

$$\#(q_n) = \frac{1}{2n+1}\binom{2n+1}{n}.$$

Each term in the product may be written as $x_{k_1} \cdots x_{k_n}$ with $k_1 \leqslant \cdots \leqslant k_n$ and $k_j \leqslant j$. Thus $\#(q_n)$ is the number of distinct sequences $(k_1, \ldots, k_n)$ of positive integers meeting these requirements.

For the remainder of the argument it will be helpful to extend $(k_1, \ldots, k_n)$ to a sequence $(k_1, \ldots, k_{n+1})$ with $k_{n+1} = n + 1$ always. We may identify each such sequence with a random

walk in the plane, beginning at $(1,1)$ and proceeding one unit to the right or upward at each step, by letting k_j denote the maximum y-coordinate attained on the line $x = j$. The number of sequences meeting the stated requirements is equal to the number of paths from $(1,1)$ to $(n+1, n+1)$ which stay on or below the line $y = x$. This number is known to be $\frac{1}{2n+1}\binom{2n+1}{n}$. (See [1], p. 73.)

[1]. W. Feller, *An Introduction to Probability Theory and Its Applications*, Wiley, New York, 1950.

Also solved by J. C. Binz (Switzerland), J. G. Brennan and M. W. Bunder (Wales), C. S. Karauppan Chetty (India), J. Dou (Spain), N. J. Fine, F. Gerrish (England), M. Hébert and J. R. Boudreau (Canada), M. Josephy (Costa Rica), O. P. Lossers (The Netherlands), J. Propp (England), P. Schumer, A. J. Schwenk, C. S. Sutton, M. Woltermann, and the proposer.

ADVANCED PROBLEMS

For instructions about submitting solutions of these Advanced Problems, which should be mailed by July 31, 1986, see the inside front cover. The solver's full post-office address should be on each sheet.

6511*. *Proposed by M. Luisa N. McAllister, Moravian College, Bethlehem, PA.*

Let $R^{m\times n}$ be the vector space of all $m \times n$ matrices with real entries. Suppose that $A = (a_{ij})$, $B = (b_{ij})$ are in $R^{m\times n}$ and satisfy $0 \leqslant a_{ij} \leqslant b_{ij}$ for all i, j. For which norms ν on $R^{m\times n}$ does the inequality

$$0 \leqslant \nu(A) \leqslant \nu(B)$$

always hold?

6512. *Proposed by Alfonso Villani, Università di Catania, Italy.*

Let (Ω, S, μ) be a positive measure space. Let A be a Borel subset of $\mathbb{C}$ (the complexes) and let $M(A)$ denote the subset of $L^1(\mu)$ defined by $M(A) = \{f \in L^1(\mu) | \mu\{f^{-1}(A)\} > 0\}$.

(a) Assume that A is a set of first category not containing zero. Prove that $M(A)$ is of first category.

(b) Assume that A is of second category. Prove that $M(A)$ is of second category.

(c) For which measures μ is the assumption "$0 \notin A$" in (a) superfluous?

6513. *Proposed by Maurice Machover, St. John's University, Jamaica, NY.*

(a) Show that the Green's function for the operator $Lu = u''$, with any boundary conditions, has no pole of order higher than three.

(b) Devise boundary conditions (the simpler the better) for which there is a third order pole.

SOLUTIONS OF ADVANCED PROBLEMS

Evaluation of an Integral

6470 [1984, 519]. *Proposed by M. L. Glasser, Clarkson College.*

Evaluate the definite integral

$$I = \int_0^\infty x^{1/2} \exp\{-a^2 x(x-b)^2/(x-1)^2\}\, dx$$

for $b \geqslant 1$.

Solution by the proposer. The substitution $x \to x^2$ yields

$$I = \int_{-\infty}^{\infty} x^2 \exp(-a^2 u^2)\, dx$$

where

$$(*) \qquad u = x(x^2 - b)/(x^2 - 1).$$

It is well known that $I = \sqrt{\pi}/(2a^3)$ when $b = 1$.

Suppose $b > 1$. Then it is readily seen from a graph of u that $(*)$ has three roots $x_1(u)$, $x_2(u)$, $x_3(u)$, each continuously differentiable and increasing for $-\infty < u < \infty$. Since x_1, x_2, x_3 satisfy the equation

$$x^3 - ux^2 - bx + u = 0,$$

it follows that

$$x_1 + x_2 + x_3 = u \quad \text{and} \quad x_1x_2 + x_2x_3 + x_3x_1 = -b.$$

Hence

$$x_1^2 + x_2^2 + x_3^2 = u^2 + 2b \quad \text{and} \quad x_1^3 + x_2^3 + x_3^3 = u^3 + 3ub - 3u.$$

Now

$$\begin{aligned} I &= \left(\int_{-\infty}^{-1} + \int_{-1}^{1} + \int_{1}^{\infty}\right) x^2 \exp(-a^2u^2)\, dx \\ &= \int_{-\infty}^{\infty}\left(x_1^2\frac{dx_1}{du} + x_2^2\frac{dx_2}{du} + x_3^2\frac{dx_3}{du}\right)\exp(-a^2u^2)\, du \\ &= \frac{1}{3}\int_{-\infty}^{\infty}\frac{d}{du}\left(x_1^3 + x_2^3 + x_3^3\right)\cdot \exp(-a^2u^2)\, du \\ &= \int_{-\infty}^{\infty}(u^2 + b - 1)\exp(-a^2u^2)\, du = \frac{\sqrt{\pi}}{a}\left(\frac{1}{2a^2} + b - 1\right). \end{aligned}$$

A Ring of Polynomials

6473 [1984, 519]. *Proposed by Mowaffaq Hajja, Yarmouk University, Irbid, Jordan.*

Let k be a field and let R be defined by

$$R = \{f(x) \in k[x] : f \text{ is monic} \quad \text{and} \quad \mathrm{GCD}(f, f') = 1\}.$$

Define the boundary operations $\odot$ and $\oplus$ on R by

$$\begin{aligned} f \odot g &= \mathrm{GCD}(f, g), \\ f \oplus g &= \mathrm{LCM}(f, g)/\mathrm{GCD}(f, g). \end{aligned}$$

Prove that $(R, \oplus, \odot)$ is a ring.

Solution by O. P. Lossers, Eindhoven University of Technology, The Netherlands. Consider the set S of all monic irreducible polynomials over k that belong to R.

Every $f \in R$ has a unique squarefree factorization into elements of S. Therefore, we may associate in a unique way to every $f \in R$ the set $\phi(f)$ of its monic irreducible factors. Actually, the mapping ϕ is a bijection from R to the collection $P^*(S)$ of all finite subsets of S. It is a matter of simple verification to show that for all $f, g \in R$

$$\begin{aligned} (f \odot g) &= \phi(f) \cap \phi(g), \\ (f \oplus g) &= \phi(f)\Delta\phi(g), \end{aligned}$$

where $\cap$ and Δ denote intersection and symmetric difference of sets, respectively. Because $(P^*(S), \Delta, \cap)$ is a ring the assertion is proved.

Also solved by Erich Badertscher (Switzerland), S. F. Barger, F. Rudolf Beyl, Stephen D. Bronn, Alberto Facchini (Italy), Luiz Henrique de Figueiredo (Brazil), Robert Gilmer, Curtis D. Herink, A. A. Jagers (The Netherlands), S. V. Kanetkar, Kenneth M. Levasseur, Jean-Marie Monier (France), A. M. Nelson (Australia), David A. Sibley, Dennis Spellman, W. B. Stewart (United Kingdom), and the proposer.

A Catalonian Sum

6480 [1984, 651]. *Proposed by Robert E. Shafer, Berkeley, CA.*

An expository article in the October 1983 issue of this MONTHLY deals with the result

$$\sum_{m=1}^{\infty}\binom{-1/2}{m}\frac{(-1)^m}{m} = \log 4.$$

Prove the following extension:

$$\sum_{m=1}^{\infty}\binom{-1/2}{m+n}\frac{(-1)^m}{m} = \binom{-1/2}{n}\left(\log 4 + \sum_{m=1}^{n}\frac{1}{m}\right).$$

Solution by D. E. Knuth, Stanford University, Stanford, California. The identity

$$\sum_{m=1}^{\infty}\binom{-1/2}{m}\frac{(-4z)^m}{m} = \sum_{m=1}^{\infty}\binom{2m}{m}\frac{z^m}{m} = 2\ln\left(\frac{1-\sqrt{1-4z}}{2z}\right)$$

is true since each expression has the same derivative and vanishes at $z = 0$. The $n = 0$ case follows by letting $z \to 1/4$. In general, set

$$\sum_{m=1}^{\infty}\binom{-1/2}{m+n}\frac{(-4z)^m}{m} = \binom{-1/2}{n}f_n(z)$$

so $f_0(1/4) = \ln 4$ and (by direct calculation)

$$f_{n+1}(z) - f_n(z) = \frac{n!n!}{(2n+1)!}\sum_{m=1}^{\infty}\binom{2m+2n}{m+n}\frac{z^m}{m+n+1} = g_n(z)$$

where (aside from the factor outside the summation) $g_n(z)$ is the generating function of the Catalan numbers beginning with the $(n+1)$st, and

$$\sqrt{1-4z} = 1 - 2z - 2zg_0(z).$$

Since

$$\frac{2z(2n+3)}{(n+1)}g_{n+1}(z) = g_n(z) - \frac{2z}{(n+1)(n+2)}$$

and $g_0(1/4) = 1$, it is easy to see by induction that $g_n(1/4) = 1/(n+1)$, and the result follows.

George Andrews introduces a new variable in a very different way:

$$\sum_{k=1}^{\infty}\frac{\left(\frac{1}{2}\right)_{n+k}(x)_k(-1)^n}{k!(n+k)!} = \frac{\left(\frac{1}{2}\right)_n(-1)^n}{n!}\left[{}_2F_1\left(\frac{1}{2}+n, x; n+1; 1\right) - 1\right]$$
$$= \binom{-\frac{1}{2}}{n}\left[\frac{\Gamma(n+1)\Gamma(\frac{1}{2}-x)}{\Gamma(\frac{1}{2})\Gamma(n+1-x)} - 1\right],$$

where the last equality is valid by the ${}_2F_1$ Gauss summation formula. The derivative of the left side at $x = 0$ is the left sum of the problem, while the ratio $R_n(x)$ of gamma products on the right

satisfies

$$R_n(x) = \frac{n}{n-x} R_{n-1}(x).$$

Thus $R'_n(0) = R'_{n-1}(0) + \frac{1}{n}$, etc. If we say Knuth works inside the circle of convergence and Andrews on its boundary, then M. R. Modak (India) has a compromise. Modak observes that

$$\sum_{m=1}^{\infty} \binom{-\frac{1}{2}}{m+n} \frac{(-1)^m}{m} = \binom{-\frac{1}{2}}{n} \frac{n+\frac{1}{2}}{n+1} \int_0^1 {}_2F_1\left(n+\tfrac{3}{2}, 1; n+2; x\right) dx,$$

where by Euler's integral representation the ${}_2F_1$ integrand is

$$(n+1)\int_0^1 (1-t)^n (1-xt)^{-n-\frac{3}{2}}\, dt.$$

The result follows by interchanging the order of integration!

Many solutions involved either some form of the integral

$$\int_0^{\pi/2} \cos^{2n}\phi \ln \sin\phi \, d\phi,$$

recurrences associated with it, or both. These were perhaps closest in spirit to the direct proof at the end of Bertram Ross' 1983 paper. A generalization involving a ${}_3F_2$ was obtained by S. K. Rangarajan (India).

Also solved by Werner P. Kohs, O. P. Lossers (The Netherlands), Syrous Marivani, William A. Newcomb, C. C. Rousseau, James C. Smith, Allen Stenger and the proposer.

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER

Calculus and Analytic Geometry (Third Edition). By Al Shenk. Scott, Foresman, Glenview, IL, 1984, xvi + 1131 pp.

Calculus with Analytic Geometry (Second Edition). By M. A. Munem and D. J. Foulis. Worth Publishers, New York, 1984, xvi + 1048 pp.

Calculus with Analytic Geometry (Second Edition). By Howard Anton. Wiley, New York, 1984, xxiv + 1108 pp.

PETER UNGAR
109 Lee Road, Scarsdale, NY 10583

The calculus industry. Our country devotes impressive resources to teaching calculus. Publishers estimate that about 400,000 students enroll in Calculus I each year. Assuming they take an average of two terms, and that college costs the student, parent and taxpayer $4,000 per year, we get an expenditure of the order of $400 million plus, very roughly, 100 million student-hours. The activity has not received the high-level scrutiny its extent would warrant.

Texts for the regular calculus sequence are more alike than those for any other course at this level. Alas, this is not an asymptotic approach to perfection. On the contrary, a prestigious group

felt three years ago that "freshman Calculus has become a rather poor course, mechanical instead of mathematical" [6]. Herbert Wilf observed that Microsoft's MuMath home computer program can do most of what is required to pass the course [11]. Clement, Lochhead and Monk [3] documented that a large part of the students pass through calculus without acquiring much if any understanding. (That we could remain ignorant or barely conscious of this for decades tells us that the nation's need for people trained in calculus is only a fraction of the present enrollment.)

Our student. The suitability of a text depends on what the student needs to learn and on what he knows to begin with. These books, like most present-day 2 or 3 term calculus courses, are based on the convenient assumption that the same calculus course can be suitable for students majoring in economics, statistics, physics, math, math education, etc. With the students' needs left so vague it is difficult to say to what extent a book meets them and I shall not even try. As to what he (or she, but I shall not constantly repeat this) brings to the course, I assume he is of average ability. He knows how to handle parentheses. He vaguely remembers rules for simplifying fractions but does not understand them well enough to know what may not be done. He remembers that $a^{1/2}$ is either $\sqrt{a}$ or $1/a^2$. He knows the quadratic formula but can not complete a square. One thing he has failed to absorb is that the letters he endlessly manipulated in algebra stand for numbers. If there are more than two or three letters in a discussion his mind fogs over, although he can cope with the same thing if one uses numbers instead. Word problems baffle him but he knows they come infrequently and, if he waits a little, they will go away. He may have memorized a few proofs years ago, in geometry. He has been free to ignore proofs since then and the whole idea is gone from his mind.

Point of view. I must warn the reader that my point of view in discussing our 3 books is one that has only limited relevance for courses taught in the usual way. I am interested in the clarity of the explanations and derivations, in exercises that require some thought and in applications in which the mathematics is used correctly and answers a sensible question.

My eyes were opened to another point of view when more than one hard-working but weak student told me they found help in a huge tome called The Calculus Problem Solver and also in the text by Leithold. These are long-winded and clumsy books. They do not offer superior instruction, even compared to the text we were using. However, these students were not looking for instruction, they were looking for instructions on how to answer exam questions. The range of questions which occur in calculus exams is quite limited. A book can show the student how to do every variant of every type of problem. To a mathematician this seems a needlessly laborious alternative to understanding the underlying principles, but many students prefer it and all our texts teach that way to some extent. For instance, Shenk has 14 differentiation formulas in a table and then repeats them all with u in place of x, and a factor du/dx. The other books tabulate only the second version, with the du/dx in every formula. No doubt this helps the students but if they understood the chain rule they would not need it.

Inappropriate rigor. All current texts attempt to be far more rigorous than is appropriate in any introductory course, especially one attended by the student I described. The proofs can not be taught to such a student; if one insists, one better not ask exam questions about them. I wondered if things were different in colleges which get much better prepared students. I found that in 4 Harvard Math 1a and 1b (Calc. I and II) final exams only about 7% of the points were given for theory questions.

Although the "rigor" in the books seems to be ignored by most instructors and nearly all students, it is by no means a harmless, inert ingredient. A student trying to read his text does not know what to ignore; he finds it unreadable. In addition, an author trying to be rigorous will not only give explanations and proofs the student can not understand, he will usually also fail to give the explanations and proofs the student could understand.

I find it difficult to visualize any reaction to these propositions other than a sleepy nod. Yet the same criticism applies to every new text publishers send me. They all seem convinced selection

committees demand rigor. Apparently, this is the advice they are getting from their reviewers. Messrs. Munem and Foulis wrote me that the original version of their book was on a more intuitive level but that this generated emphatic protests from nearly all reviewers. In order to help change the minds of such colleagues, I would like to devote a page or two to point out that the present approach is gravely defective, even from a logical point of view, in addition to being pedagogically inappropriate.

Let us look at the predicament of a student who knows his high school algebra and wants to learn what a proof is. He is not allowed to take for granted everything that is obvious to him. What, then, can he take for granted and what steps are permissible in drawing conclusions? The explicit answers he gets to these questions are only fragmentary, so he has to get a feeling for it by observing how his books proceed. This would in itself put an unreasonable burden on the student but to make matters worse, his books are quite confused about what constitutes a mathematically rigorous argument.

Take the 2nd year algebra book [**5**] as an example. This is one of the better books; its senior author was specially honored by our Association. On p. 21 the student is told that because any pair of numbers has exactly one sum, $a = b$ implies $a + c = b + c$. The student will be shocked that even an implication as obvious as the above needs justification and he will be bewildered by the one given. He may wonder how one justifies "$a = b$ implies $\{a, c\} = \{b, c\}$", since the expressions in the second equation do not denote exactly one real number.

At this stage the student must feel tied down like Gulliver by countless invisible ropes. Paralysis will give way to utter confusion later in the book, when i is introduced thus: "i is a solution of $x^2 + 1 = 0$". That a thing with such an unlikely property was just waiting in the shadows to be summoned onto the stage of Mathematics should apparently be more obvious to him than the simple implication on p. 21, since on this occasion no further justification is given. What else is lurking in the shadows that one may introduce without any apology? A solution of $x + 1 = x$ perhaps? And which of the two solutions $x^2 + 1 = 0$ supposedly has is i and which is $-i$?

By the time the student gets into calculus complex numbers have been forgotten, except by Shenk who uses complex exponentials discussing vibrations near the end. Nothing outlandish is defined into existence here; rather, the opposite occurs if you are an Anton student. He explains, clearly enough, why the area between two curves is $\int(f - g)\,dx$, but fearing he was not rigorous enough, he calls the formula a definition of area, ignoring that area has already been defined in grade school by counting squares, and that he had been discussing areas long before this definition.

Next comes the formula for the area as a y-integral. How can one be rigorous without proving it? By calling it a definition. They don't prove definitions, do they? (Later, between the 2nd and 3rd definition of volume, a reviewer seems to have waked up and objected. We find there a belated note that if one gives two definitions of the same thing one really ought to show they are equivalent.)

Shenk, too, is defensive about justifications for the formulas for areas and volumes but knows better than to call all of them definitions. He calls them rules. They don't prove rules, do they? Or do they?

All three books derive the limit $(\sin\theta)/\theta \to 1$ as $\theta \to 0$ (1) in the usual way, by first establishing $\sin\theta < \theta < \tan\theta$ using areas and then after a little algebra the squeeze law gives (1). This is a nice, ingenious argument but it is an illusion to think it is more rigorous than saying that Fig. 1 makes it obvious that for small θ, $(\sin\theta)/\theta \approx 1$. For one thing, areas of regions bounded by curves have not been treated rigorously. Arclength of a circle also enters because we are using radians. If arclength was defined at all, it was using polygons with small sides. This amounts to postulating that for a short arc of a circle, chord length/arc length ≈ 1. But that is what we are attempting to prove. ($2\sin\theta$ is the chord of 2θ.)

Students enter Calculus with a shaky grasp of fractional exponents and logarithms. We ought to reinforce their previous knowledge of these, rather than confuse them with the new and

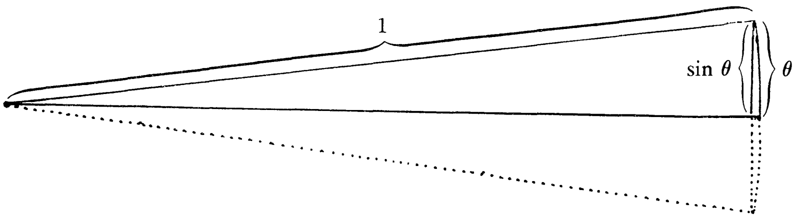


FIG. 1

circuitous approach of defining the logarithm as an integral and the exponential function, which is the easier function to understand, as the inverse of the logarithm. I particularly regret that the basic formula $da^x/dx = \text{const} \cdot a^x$, which is an almost immediate consequence of the laws of exponents, appears at the end of an argument involving logs and inversion of functions. The reason for introducing logs and exponentials in this manner is that it makes it easier to define powers with irrational exponents, and to prove that the exponential function is differentiable.

The pursuit of rigor is clearly in conflict with sound pedagogy here. Most students would not worry about irrational exponents and the few with such a highly developed sense of logic could easily figure out what they mean. Moreover, it is an illusion that we have sacrificed sound pedagogy in the interest of a superior logical structure. Anyone with a sense of logic developed enough to worry about powers with irrational exponents would also have worried about how one performs the four simpler arithmetic operations with irrational numbers, and none of these books bothers to say anything about that.

Anton is especially eager to discuss irrational numbers, so let me point out that introductory calculus is a course in applied mathematics, and there are no irrational numbers in economics or biology or even in physics. Take r = proton mass/electron mass = 1836.10... for example. A definition of r refers to measurements which can not be carried out with perfect accuracy in this noisy universe. So the 100th digit of r is not only unmeasurable, it is undefinable. The proposition "r is irrational" is therefore meaningless; it may be rational or irrational in a particular theory, just as the number of people in the U.S. at a particular moment might be irrational in some model, but that is just an artifact and examining artifacts in minute detail does not constitute rigorous thinking.

Limits and infinitesimals. Endless discussions of limits seem obligatory in calculus texts and are a major reason for their unreadability. Before any derivative is computed, most books have a section on the limit of a function at a point. Students find this boring and difficult because it seems pointless. I can say on the authority of Professor Serge Lang that it is pointless [**8**]. Going to the limit while finding tangents or in other meaningful contexts causes no great difficulty, especially since the limits which occur naturally, such as $\lim_{h \to 0} 2x + h$, can usually be read off at a glance. There is no need for a preparatory discussion or to belabor such points as $(\lim u)(\lim v) = \lim uv$. Those who want to teach ε and δ can do so later in the more natural context of convergent sequences obtained when solving equations numerically, where the limit can not be read off.

The chapter on applications of integrals is a morass of limits in most calculus books. Volumes by disks, volumes by cylindrical shells, arclengths, surface areas, etc., all become long stories, starting with sums and going to the limit. This is boring and it slows down and inhibits the student. Having obtained the differentiation rules and the definite integral by going to the limit, one should consider dx and dy to be infinitely small increments, with the understanding that the argument can be made more precise by starting with finite increments and going to the limit.

The unwillingness to use infinitesimals in this sense is the most devastating consequence of the striving for rigor in calculus texts. Of all the current texts I know, only Munem & Foulis (M & F, from now on) and Campbell & Dierker [**2**] use infinitesimals. As M & F point out, the student has to think of dx and dy as small increments if he wants to understand how to formulate a problem

in physics or some other field in terms of calculus. Try to derive the identities of thermodynamics any other way. (None of these books ventures into that but a competing book struggles for half a page with a formula derivable in one line using infinitesimals.) On a more elementary level, the majority of students, enjoined not to think in terms of small increments, can not understand how to set up the simplest flushing of a tank or change of population problems; fortunately it suffices to memorize the equations or even just the solutions to pass their tests. Even the problem of finding the rate at which the water level in a tank is rising becomes needlessly complicated. All three books tell the student to write down $V(h)$ and differentiate it, and if $V(h)$ is too complicated they give him the formula. Anton, for example, tells his students to work from the volume of a truncated cone, $h(r_1^2 + r_1r_2 + r_2^2)/3$, to solve a problem in which the water surface is a circle of radius 3. Yet all the student needs is that $dV = A\,dh$. (M & F later tell him that.)

But, our purist critics will say, you can't say what an infinitely small increment is so how can you talk about it? This, again, ignores that introductory calculus is an applied course and that the students do not want and can not handle a high degree of rigor.

While one can not say in the abstract what is infinitely small, in a given context there is no difficulty. For instance, if we have a piston compressing a gas, pV^γ = const., then 1 micron or less will certainly be infinitely small. The purist will object that this is inadmissible, since the 5th significant digit of dp/dx will depend on whether dx is 1 micron or .1 micron, making dp/dx ambiguous. That is true but it does not matter because in fact pV^γ is not exactly constant for any value of γ, so the 5th significant digit we may compute from our equations is meaningless anyway.

The mathematician who thinks he is furthering the cause of precise thinking when he makes a fuss about the derivative being a limit and not the ratio of two small increments is mistaken. Populations and even quantities of water change by discrete units so the functions which really represent them do not have derivatives. Even the definition of velocity as a limit is meaningless in physics; Heisenberg taught us that a particle does not have a precise location at every instant. The reason for using derivatives and integrals instead of difference quotients and sums in these contexts is that the mathematics is much simpler that way. So we go to the limit in such problems for mathematical convenience, and it is not an offense against logic to stop a little short of the limit when that makes things easier.

M & F try to placate critics by saying that Abraham Robinson's nonstandard real number systems justify the engineer's and physicists' use of infinitesimals. This is not so; in Robinson's non-standard number systems infinitesimals can be interpreted as numbers but the physicists continue to use standard real and complex numbers, and operate with infinitesimals as if they would denote small but standard numbers. Robinson's work ensures the consistency of algebraic manipulation of infinitesimals but it does not justify treating dx and dy as small but ordinary real numbers in problems of physics, etc. The much simpler considerations we just outlined suffice for that purpose. (There exists an introductory calculus book which is based on nonstandard number systems. Basing the theory on number systems different from the one used in all applications, and impossible for the student to understand, can only be described as educational malpractice.)

It is bad enough that Anton and Shenk do not encourage students to think of dx and dy as small increments. It is worse still that they, as well as all other books on the market, including M & F, tell the student that dx is a new independent variable. Dr. Alonzo Church wrote: "If dx is a new independent variable we may call it z. Since $\int x\,dx = x^2 + C$, $\int xz = x^2 + C$. But what kind of operation could $\int$ be to yield $x^2 + C$ when applied to the symmetric expression xz?" [**1**]. (In M & F, dx is a new independent variable in the theory and an infinitely small increment in the applications. This must be confusing.)

So much for the pedagogical and logical weaknesses of the attempts at rigor in our textbooks. I should hasten to state that I am not advocating "cookbook" courses. Elementary calculus is often taught as if differentiating formulas were a marketable skill, in contrast to knowing the theory, and students are receptive to this notion. In fact, being able to differentiate x^4 is no more marketable than knowing why $(uv)' = u'v + uv'$; if we want to teach a useful skill of the cookbook type we should teach how to make beef strogonoff.

A math course should impart understanding, not just mechanical skills. It should show how one can solve problems by reasoning. Fortunately, practically everything in our present calculus course (even the error formula for Simpson's rule—see the instructive volume [**10**]) can be demonstrated very simply. Often a diagram says it all, as above. Rigorous proofs must not depend on diagrams but it is misguided to apply that restriction to a course where the aim is to make the student see why the propositions are true, not to minimize the number of facts accepted without proof.

Calculus and the computer revolution. The effect calculators and computers had on our three books is negligible. A student without a calculator would barely be inconvenienced. The reason for this must be poverty of ideas, since even a new programmable calculator costs less than a calculus book. Anton and M & F each devote a total of a dozen pages to numerical methods: Newton, trapezoidal and Simpson's rule. Anton also has his own, needlessly laborious method of interval halving. Shenk is somewhat better. In addition to the above topics, he introduces the student to numerical solution of differential equations and briefly mentions the method of iteration for solving equations (with 1 unknown). It is sobering to note that even he says less about this simple and immensely fruitful idea than Legendre does in his Théorie des Nombres (1830).

Shenk is the only one of our books with a topic truly of the calculator age: 13 pages on the difficult subject of Newton's method in 2 variables. Unfortunately the student is not given meaningful problems that would justify the labor. Locating an approximate solution is far from easy in two variables; Shenk simply gives starting values to the student. In this he is needlessly cautious; I tried various initial values for some of his examples and divergence was exceptional.

In none of these books are numerical methods integrated into the course; outside their special sections problems solvable without them are the rule. Shenk has more exceptions to this rule than the others, but only a handful. The scarcity of genuine calculator problems is an indication of the difficulty of modernizing this course. Curiously, additional evidence of how hard it is to get into new habits of thought is provided by one of Shenk's best calculator problems. The student is asked to determine points on curves such that the sum of the distances from two given points is least. He is told to differentiate the sum, set the result equal to 0 and solve the resulting equation numerically. This is quite a mess and can not be done on the simplest programmable calculators, as the student is warned (p. 261). Yet all one has to do is to program the formula for the sum of the distances; there is no need to compute the derivative because one can find extrema by interval halving type methods as easily as one can solve equations.

Simpson's rule is not a new or difficult topic, but a surprising number of authors have no feeling for how accurate it is when properly applied or when not to apply it. Anton asks the student to compute

$$\pi = 4\int_0^1 1/(1 + x^2)\, dx$$

using Simpson's rule with $n = 10$ but instructs him to round all computations to 4 digits. If he does that he gets no more accuracy than he would with $n = 4$ and he is deprived of the pleasure of having computed π to 8 significant digits.

Let

$$I(a) = \int_0^a (1 - x^2)^{1/2}\, dx.$$

M & F's students compute $\pi = 4I(1)$ using $n = 10$ and get a miserable 3.127. Shenk also assigns this computation but he next assigns finding π using $\pi = 8I(1/\sqrt{2}) - 2$, again with $n = 10$, which gives 3.141580. I suppose Shenk wants the student to observe that Simpson's rule is not good if the curve has a vertical tangent since no quadratic function can approximate such a curve well, but that the rule is very good for more regular functions. However, it would be better to point these things out since most students are not likely to observe them on their own.

The results of calculations are affected by errors in the data and by roundoff errors. The latter topic is completely ignored by all three books, even though they all review inequalities and absolute values in the first chapter (Anton more thoroughly than the others) and this would be a source of worthwhile problems where they all have difficulty coming up with some.

Graphics. The easiest way to make a thing of some beauty is to draw an accurate graph of a simple equation. Many students will never be aware of this because almost all the curves in their books are fakes. Our three are not bad in this respect. Shenk has many computer-drawn curves and M & F have some too. The computer's contribution to Anton's curve collection is a few zigzags, put in perhaps to show he is one of the boys. The boys could have shown him how to do better even with a simple matrix printer. The ugly fakes, which occur in all three books, do have some pedagogical value. One can ask which law of geometry is most visibly violated. For instance, the curvature of Anton's Archimedean spiral is not monotone; M & F have a hyperbola with a point 3 times farther from the directrix than from the focus.

All three books have graphs of specific functions with no units marked. Graphing is one thing they do in the real world too, and the boss will certainly not accept graphs without scales.

Shenk was the first author to use computer-generated images of graphs and level curves of $f(x,y)$. Amazingly, his 1977 1st edition has not yet been matched in this respect by any other book. These graphics and the many authentic real-life 2-variable functions he gathered as examples make the chapter on partial derivatives in all the other books look pale by comparison.

The other two books also show some computer-drawn surfaces. You would expect that in their chapters on lines and planes all our books would discuss orthogonal and central projections, basic to computer graphics and in fact the most important topic in geometry even before computers. None of them do. Worse, none of them seem to know how to draw the simple 3-dimensional diagrams they need, although a computer with any kind of graphics capability or even a calculator and graph paper make it painless.

Our authors are just totally ignorant of perspective. For instance, they all think one may prescribe any 3 vectors in the plane as the images of i, j, k and the linear mapping this generates will be a suitable projection. Such a mapping will indeed be equivalent to a parallel projection plus a change in size, but the projection need not be orthogonal and will produce a distorted image. A sphere, for instance, would project into an ellipse. The image plane is tilted by up to 26° in Anton and even more in others. To cite another example, if you draw a bare cone and three coordinate axes and take the precaution of not marking units there are not many laws of perspective you can violate—the angle of tilt can not be determined then—but all three of our books manage to violate some. This helplessness in trying to depict the simplest objects in space contrasts strikingly with the books' beautiful design and lavish production. The same comments apply to every calculus book I examined recently, 9 in all. This is very embarrassing for American education, although foreign books can be just as bad [**6a**].

The choice. I should keep in mind that most colleagues read calculus book reviews to find out which book will make teaching the usual course easiest; reforming it is not one of their alternatives or priorities. The choice is Anton, and not just among these three. The book is very carefully written by someone who seems to be in close touch with the average student at the average college. Just a small example: most books discuss convergence of the sequence $a_1, a_2, \ldots$ and then convergence of the series $a_1 + a_2 + \cdots$ and students tend to confuse the two concepts. Anton helps to avoid confusion by calling the series $u_1 + u_2 + \cdots$.

I made detailed comparisons of many sections and Anton's presentation is clearest in most cases.

The one major exception among the basic topics is Anton's introduction of the $\int f(x)\,dx$ symbol before the definite integral and the fundamental theorem. In this at least most recent texts, including Shenk, dared to deviate from Thomas. The definite integral symbol is very suggestive, and once we have it, the indefinite integral symbol appears fairly natural as a sloppy form of it. Of course, from a "practical" point of view, the student needs the symbol $\int f(x)\,dx$ only to guide him

when integrating by substitution and it will perform that service even if he has no idea why the antiderivative is denoted this way.

Differential equations are presented best by Shenk. His computer graphics enable him to show the student the direction fields associated with various equations. There are books which don't even tell the student that a differential equation is a direction field. M & F, like Thomas, introduce separable equations 13 pages after the indefinite integral (and before the definite integral), and follow this immediately with a long discussion of the simple harmonic equation. I doubt the students are ready for these topics so early. Anton's usual light touch is conspicuously absent when he starts the discussion of 1st order linear equations with the most general case, and gets to $y' = ky$ at the end.

Problems. A few remarks about the problems in our books. Shenk has a unique feature: he eases the transition from examples to problems by having several "study problems" first, whose solutions are given in the text but in a shorter form than the solutions of the examples. He also gives answers and sometimes even brief solutions in the back for some of the regular problems, which he marks. The other two books use the rigid procedure of giving answers to odd numbered problems.

Competitors call Anton babyish. It is clear that leaving out things which students find difficult contributes much to his success; just compare his section on graphing to other books. However, he does have problems which are not mere chores. There are not many but I can not fault him for that because such problems are hard to come by. (The text with the best collection is Sherman K. Stein's, a book also notable for its unhurried style and excellent explanations.)

Among our 3 books Shenk has the most challenging non-routine problems but a little cynicism is in order when one sees such problems. I used to be impressed that colleagues teaching in colleges no more selective than mine could assign a problem such as: "Suppose that of P animals, $s(t)P$ are still alive t years later. Show that if $P(t)$ is the population and $r(t)$ is the birthrate at time t, then

$$P(t) = P(0)s(t) + \int_0^t s(t-u)r(u)\,du." \qquad \text{(p. 415)}$$

Note that a factor $P(u)$ is missing under the integral sign and this would cause great unhappiness among students trying to do the problem. Since it has been missing since the first edition (p. 281), it is reasonable to conclude that neither the author nor his 37 reviewers ever assigned this problem, or the one just preceding it which is also flawed and unchanged from the first edition.

M & F also have a much better than average collection of problems that are not mere chores, in addition to masses of the latter. I have not noticed any that one could not assign to the better students.

Applications. Applications are in vogue now. Anton says in his preface: "An abundance of applications to physics, engineering, biology, population growth, chemistry and economics appear throughout the text". Fortunately for authors, a book is not a can of soup and there is no penalty for misrepresenting its contents. In fact Anton keeps applications, and all the frustration they bring to the classroom, to a minimum. For instance, there are none in the 72-page stretch pp. 467–538. (Unsurprisingly, this coincides with a traditional high point of mindlessness in Thomas-style calculus books, "techniques of integration". For a shorter and more intelligent presentation of this material as well as other good ideas, see [**9**].) I should add that no one could deliver on Anton's promise by the usual procedure of starting with the mathematics in Thomas and straining to find applications. (I do not mean to imply that Anton strained himself in this regard.) One could provide a worthwhile application for everything by the easier method of leaving out what is useless. That would also provide space for new material.

Both Shenk and M & F are among the most seriously applications-minded books on the

market. Shenk presents far more authentic data than any other calculus book, e.g., a graph showing the amount by which load x compresses a spinal disk, or the amount of solar energy per day reaching the top of the atmosphere as a function of latitude and date. These data are interesting in their own right but it must be admitted that the rate of change or whatever the student is asked to compute from them often provides little additional insight.

M & F have the most applications in the usual sense. They don't just throw them in, they back them up with many exercises. Reading the book, this hard-to-please reviewer pondered where the line should be drawn between applications and taking over chunks of mechanics and economics courses. These departments may claim not only that their students' time is wasted if they are taught in Calculus what must also be taught in these other courses, but also that they can explain consumer's surplus, conservation of energy or Kepler's laws better than M & F or Shenk. They may well be right. For instance, M & F use so many letters when they present consumer's surplus that the student's mind is sure to fog over. To make matters worse, they throw integration by parts, which the student has not seen before, into the middle of it. Shenk gives consumer's surplus as an assignment with a brief solution in the back. Anton does not discuss consumer's surplus.

In 1947 C. C. MacDuffee mused about a combination course in calculus and physics [**1**]. He thought there might be a lack of instructors competent in both subjects. I am reminded of this when Shenk says the curve of quickest descent from A to B is half an arch of a cycloid, and draws a picture to show he really means it. Not one of the 37 math professors who reviewed his book knew this classical problem well enough to correct him or even to notice that there is no such half-arch unless the angle between AB and the vertical is arc tan $\frac{1}{2}\pi$. Anton put the same thing in his book. His 2nd edition had 22 reviewers. Another item in Shenk which will raise eyebrows in the physics department is Study Problem 8 on p. 462 which is about the motion of a shadow in sunlight. The problem can not be solved without knowing the date and the latitude, which are not given. Reading the solution one discovers that to Professor Shenk and his 37 reviewers it is a matter of course to assume that the sun moves in a vertical plane. Anton has 5 pages on the force on a submerged surface without any indication in the text or the diagrams that the force is a vector and that it is perpendicular to the surface element. He then assigns a problem to compute the total force on a curved surface. Incidentally, in both Anton and Shenk the density ρ is the weight per unit volume. I should say that the few slips in Shenk and Anton can not be compared to the ignorance of physics shown in certain competing texts. Of course, if the authors of those had stuck to a few old staples as Anton did, they would not have erred either. But they could also have learned their physics, as M & F did.

History. Both Shenk and Anton enliven their books with many pages of history. We should not imagine the history of math will interest students who are bored with math itself. (Even the minority who like math tend to prefer learning new math to old.) However, I think it is valuable to learn about the circuitous paths through which the human race arrived at its insights. Shenk gives much real enlightenment, presenting pre-calculus area determinations for instance. By contrast, Anton concentrates on biographical details, with a weakness for relaying admittedly apocryphal stories.

Summation. All three books have something distinctive to offer. Anton is very carefully written and in addition he shields the student from difficult material more than any competing book. Shenk stimulates with exceptionally intelligent presentations in many places and disappoints with careless ones in others. M & F teach the eager student not just the theory but also lots of applications, presented with a variable degree of skill and reinforced with lots of good problems.

Will an up-to-date calculus course emerge through gradual evolution? With all due respect to the efforts of these authors and the many others whose books I have looked at over the last 10 years, the answer seems to be no. Major reform is needed. In the sixties math courses created by committees of eminent people turned out to be disappointing but the lessons of that experience should make it possible to do it better now.

References

1. Tom M. Apostol et al., editors, Selected Papers in Calculus, Mathematical Association of America, Washington, DC, 1969.
2. Howard E. Campbell and Paul F. Dierker, Calculus, 3rd ed., Pringle, 1982.
3. John Clement et al., Translation difficulties, etc., this MONTHLY, 88 (1981) 286.
4. The College Mathematics Journal, Nov. 1984, "Forum".
5. Mary P. Dolciani et al., Algebra & Trigonometry 2, Houghton Mifflin, 1971.
6. Focus (Newsletter of the Mathematical Association of America), Jan. 1983, 2.
6a. Branko Grünbaum, Geometry strikes again, Math. Mag., Jan. 1985, 12.
7. Peter Hilton, Review of "The Future of College Mathematics," this MONTHLY, 91 (1984) 452.
8. Serge Lang, First Course in Calculus, preface.
9. Jerrold E. Marsden and Alan D. Weinstein, Calculus, 2nd ed., Springer-Verlag, New York, 1985.
10. David A. Smith, Interface, Calculus and the Computer, Houghton Mifflin, 1976.
11. Herbert S. Wilf, The disk with a college education, this MONTHLY, 89 (1982) 4.

Complex Analysis, a Functional Analysis Approach. By D. H. Luecking and L. A. Rubel. Springer-Verlag, 1984. vi + 176 pp.

LAWRENCE NARICI
Department of Mathematics, St. John's University, Jamaica, NY 11439

The expression "functional analysis" was coined to describe the investigations of certain collections of functions—continuous functions, functions of bounded variation, holomorphic functions, for example—as vector spaces in which limits could be considered and analysis performed. A typical one was the space $C[0,1]$ of continuous functions on $[0,1]$ topologized by the norm

$$\|x\| = \sup\{|x(t)| : t \in [0,1]\}, \qquad x \in C[0,1].$$

There were sufficiently many of these spaces to warrant a general investigation of normed spaces and, later, of vector spaces topologized differently, of topological vector spaces (TVS's) that is. The revelations of the more general approach are spectacularly enlightening at times. A result such as Weierstrass's theorem on the uniform approximation of continuous functions by polynomials ceases to seem a peculiarity of polynomials and is revealed as a special instance of a more general phenomenon, suggesting many new possibilities. Indeed, there are many areas of classical analysis which may be brightened by the light of functional analysis. The book under review, for example, investigates complex analysis using functional analytic methods and there is also Nachbin's recent little gem [**1**] introducing the calculus this way. I do not mean to argue, incidentally, that all of analysis can or should be done in this manner, as I think that the generally more constructive approaches of classical analysis have their own secrets to reveal. I mean only that the functional analytic perspective provides an enlightening alternate approach. One *should* see Weierstrass's continuous, non-differentiable function as well as the category argument that such functions exist as the scenery along each route makes both journeys worthwhile. To go by plane or by car? The plane gets you there faster (and functional analysis is the plane) but you don't see as much along the way.

The preferred form of the Axiom of Choice for the functional analyst is the Hahn-Banach theorem. Its assertion of existence is that, on most TVS's, a continuous linear form (or 'functional') defined on a linear subspace has a continuous linear extension to the whole space. A consequence of the theorem is the following criterion for density:

(D) If the only continuous linear form which vanishes on a linear subspace M is 0, then M must be dense in the ambient TVS.

The *continuous dual* (or just *dual*) of a TVS X is the vector space X' of all continuous linear

forms on X. In many cases the dual may be neatly characterized. If f is a continuous linear form on euclidean n-space $\mathbf{R}^n$, then there is a unique n-tuple (b_i) such that, for any (a_i) in $\mathbf{R}^n$, $f((a_i)) = \Sigma a_i b_i$. Thus, the dual of $\mathbf{R}^n$ "is" just $\mathbf{R}^n$ again. As the space increases in size (i.e., dimension) and, consequently, complexity, the depth and difficulty of characterizing the dual grows concomitantly. If X is the Banach space $C[0,1]$ mentioned earlier, a deep theorem of Riesz shows that the continuous linear forms f on X are all of the form

$$f(\cdot) = \int_0^1 (\cdot)\, dy(t),$$

where y is a function of bounded variation on $[0,1]$. Thus, the dual of $C[0,1]$ is essentially the functions of bounded variation on $[0,1]$.

The ambient TVS for the results in "Complex Analysis" (the authors call it "our hero") is the (linear) space $H(G)$ of holomorphic functions on an open subset G of the complex plane. $H(G)$ carries the *compact-open topology*, the weakest topology on $H(G)$ that makes the maps

$$p_K(x) = \sup\{\,|x(t)|: t \in K\}, \qquad x \in H(G),$$

continuous as K runs through the compact subsets of G. As a sequence (x_n) of members of $H(G)$ converges to x in $H(G)$ in this topology if and only if $x_n \to x$ uniformly on each compact subset K of G, it is also called the "topology of uniform convergence on compact sets." Subsets of $H(G)$, incidentally, are compact in the compact-open topology if and only if they are closed and bounded. The dual of $H(G)$ is essentially the space of holomorphic functions on the complement of G. If G is the open unit disk D, for example, then each continuous linear form f on $H(D)$ is of the form

$$(*) \qquad f(\cdot) = (1/2\pi i)\int_C (\cdot)\, y(z)\, dz,$$

where C is a circle in D and y is holomorphic on and outside of C.

The density criterion (D) and the characterization $(*)$ of the dual of $H(G)$ are the principal functional analytic tools used in "Complex Analysis." Their first use is in proving Runge's Theorem (a simple form of which asserts that the polynomials are dense in $H(G)$ when G is simply connected) from which follow a strong form of the Cauchy Integral Theorem and Rouché's Theorem. The Mittag-Leffler Theorem (on the existence of meromorphic functions with specified poles and principal parts) is proved first by constructive methods and then, for contrast, by functional analytic methods. Some deep interpolations theorems are also proved.

I admire the plan of "Complex Analysis," but certain things mar its execution for me: the lack of an index, for one, the instances where proofs are only "sketched," for another. But, for someone with some background in functional analysis, I think it constitutes an interesting second look at complex analysis.

Reference

1. Introduction to Functional Analysis: Banach Spaces and Differential Calculus, by Leopoldo Nachbin, Monographs and Textbooks in Pure and Applied Mathematics #60, Marcel Dekker, New York, 1981.

Intermediate Real Analysis. By Emanuel Fischer. Springer-Verlag, New York, 1982 xiv + 770 pp.

Grant V. Welland
Department of Mathematics, University of Missouri, St. Louis, MO 63121

When I was a graduate student, an approach to reading a book in mathematics came to my attention; it is an approach that served me well whenever I chose to apply it. The approach is

based on the hope that every author writes a book to address an important problem, with the concept of a problem widely interpreted. One should discover the nature of the problem before seriously reading the book. This is an amplification of an idea that was expressed by my teacher in a beginning course in analysis: he urged us to read the statement of a theorem and then prove it by ourselves without reading the proof in the text until we were stuck.

In reading a mystery novel, I recommend that one resist the temptation to look ahead in order to enjoy the story as the author carries us to the final solution. But in reading a book in mathematics I encourage everyone to read enough of the book, in helter-skelter fashion if necessary, to gain a firm grasp or at least an intuitive understanding of the problem being treated. After this has been accomplished, the reader may be willing to move on to seek solutions.

In this regard, a pedagogically motivated book has a difficult time defining its purpose. It must mix contents with methods, addressing separate problems of essentially distinct types and developing solutions to these within the body of one work. The result is often a product of uneven mix. I have always preferred a textbook that displays a clear and direct purpose. A good example of this is Gelfand's book, *Lectures on Linear Algebra*, for which the author appears to have had a homogeneously talented audience and a singleness of purpose.

I used these ideas to guide me in studying the text to be reviewed here, but it was a frustrating exercise. A textbook must present a course to a classroom of varied majors with a wide range of talents and interest. Only the most skillful and conscientious teacher has a chance to chart a course suitable for all. There have to be choices, matters of style, and conflicts of personality and preference; and the questions of talents, motivation, and interests of these students cannot be ignored. I believe that the purpose of this book is to teach young maturing students the foundations and principles of analysis. There is no single, important problem addressed here.

For comparison, I looked for a book to use as a foil. In his preface, E. Fischer writes that his book is supposed to fill a gap that is not filled adequately by those that already exist. I selected one, Walter Rudin's book, *Principles of Mathematical Analysis*. It certainly has enjoyed a good share of success, and both books claim to target a similar audience. By comparison one finds Rudin's book is just greater than a third the size of Fischer's. Perhaps this is a cheap comparison: No one would deny the value of a book because it is fat. (Among my favorites is Courant's *Differential and Integral Calculus*.) But any student who has to buy this book is going to notice the difference in an essential way. Most of the extra material in Fischer's book is there to deal with a pedagogical problem he has expressed in the preface. He contends that there are not enough meaningful or motivating examples in most books to illustrate the theory and principles of analysis for a beginning student. This echoes the plea for course relevance that was so common in the sixties and seventies. He responds with great effort and the inclusion of much additional material.

Understanding certain topics, the definition of a limit for example, does not improve with elaborate examples. For most beginning students some easy examples are sufficient and require time and thought before the basic idea behind them is fully grasped. In fact, it is pedagogically sound to restrict the application of principles and theory to simple examples while the beginner is still developing his skill and understanding. Otherwise, one is likely to divert his attention from the principles at hand. My advice is to leave the elaborate examples for additional courses, and do a good enough job so that the student understands the need for the techniques and will continue his studies. With these comments, I find myself staring face to face with Fischer contradicting his statement, "...good pedagogy in mathematics should give substance to abstract and general results by demonstrating their power." I don't really disagree with him. I can only respond, "All things in their time."

On the positive side, I did find much of the additional material pleasing. Many of the examples and applications are selected from topics in special functions and orthogonal polynomials. (The author warns us in advance of his love for this material by dedicating the book to Wilhelm Magnus.) For the teacher who occasionally has the time to treat special examples in his course, there are many good ones in this book. I also should add that some problems on Putnam exams

can be treated using the material found here.

Books in analysis tell us something about analysis, just as the statement of a theorem gives us information about a subject area. Using Rudin's book, we find principles above all else and perhaps a beginning reader could be misled to view analysis as too theoretical and limited. In Fischer's book we take a peek past the principles at the astronomically vast accomplishments of analysis. I don't think this is bad; it just seems to be too much for one book and certainly too much for one course. It would have been better if he had set out to write a book of examples, the ones he clearly loves, in the spirit of the book by Gelbaum and Olmstead (*Counterexamples in Analysis*).

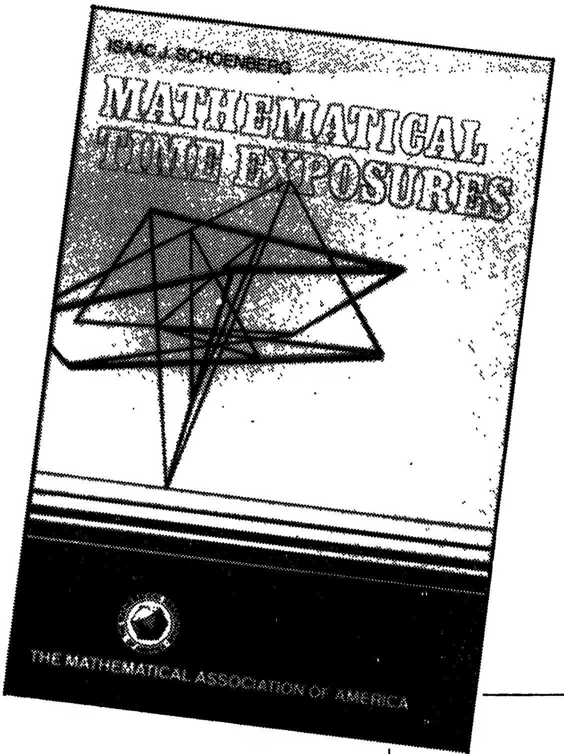
ACKNOWLEDGMENT

The editors thank the following persons who have refereed manuscripts between mid-1984 and mid-1985:

Harvey L. Abbott, James C. Abbott, Alexander Abian, Janos Aczel, Alfred Aeppli, David J. Aldous, Herbert Alexander, George E. Andrews, Tom C. Apostol, Barry C. Arnold, Richard A. Askey, C. E. Aull, David Avis, Edward A. Azoff, B. J. Ball, Edward J. Barbeau, Jr., Jeff Barnett, David W. Barnette, Robert G. Bartle, Estelle Basor, Paul T. Bateman, James E. Baumgartner, John V. Baxley, J. Thomas Beale, David J. Bell, John J. Benedetto, David M. Berman, Gerald Berman, Bruce C. Berndt, Richard A. Bieberich, Patrick P. Billingsley, Paul Binding, Richard L. Bishop, R. G. Bland, Andreas A. Blass, William J. Blundon, Colin R. Blyth, T. S. Blyth, J. A. Bondy, Louis R. Bragg, Fred Brauer, Joseph P. Brennan, David M. Bressoud, Robert H. Breusch, Robert C. Brigham, Duane M. Broline, Ezra Brown, Richard A. Brualdi, Paul S. Bruckman, Andrew Bruckner, Edwin O. Buchman, R. Creighton Buck, Robert Bumcroft, Donald L. Burholder, Owen Burkinshaw, James Callahan, Douglas M. Campbell, Stephen L. Campbell, E. Rodney Canfield, Frank B. Cannonito, Robert S. Cantrell, Gerald T. Cargo, Bille C. Carlson, Alfred S. Cavaretta, Jr., Robert Chaffer, C. Y. Chao, Paul R. Chernoff, Lindsay N. Childs, Tseng Yeh Chow, Charles K. Chui, Fan R. K. Chung, K. L. Chung, Joseph A. Cima, David W. Cohen, Paul M. Cohn, Charles C. Conley, J. Brian Conrey, Kenneth L. Cooke, Roger Lee Cooke, John Corcoran, Clifton Corzatt, Carl C. Cowen, Henry H. Crapo, F. H. Croom, James T. Cross, Marc Culler, Robert J. Currier, Charles W. Curtis, Morton L. Curtis, Thomas W. Cusick, C. DaFermos, George B. Dantzig, Richard B. Darst, John Daughtry, Chandler Davis, P. J. Davis, A. Michael Dawes, Jane M. Day, Richard A. Dean, Herbert A. DeKleine, Robert L. Devaney, Seymour Z. Ditor, Robert S. Doran, Daniel J. Dougherty, Jr., Randall Dougherty, Steven Dragosh, David Drasin, Vladimir Drobot, L. E. Dubins, Richard M. Dudley, James Dudziak, Richard J. Duffin, Douglas Dunham, Charles F. Dunkl, William F. Eberlein, Charles C. Edmunds, Charles H. Edwards, Jr., Harold M. Edwards, Paul Ehrlich, Susanna Epp, James Epperson, Paul Erdös, Garrett J. Etgen, Stewart N. Ethier, Ronald J. Evans, Graeme Fairweather, H. K. Farahat, R. Fefferman, Lawrence A. Fialkow, Michael A. Filaseta, Ross L. Finney, William Firey, Peter C. Fishburn, Evan Fisher, Stephen D. Fisher, Harley Flanders, Martin E. Flashman, R. Fletcher, Joel N. Franklin, John M. Franks, D. A. S. Fraser, John A. Fridy, Laszlo Fuchs, Lisl Gaal, Michael Gage, David Gale, Joseph Gallian, Steve Galovich, Martin Gardner, Bernard R. Gelbaum, Ira M. Gessel, William B. Giler, John S. Giles, Robert Gilmer, M. Lawrence Glasser, Jerome A. Goldstein, Daniel A. Goldston, Solomon Golomb, Victor Goodman, Basil Gordon, Ronald J. Gould, Sandy Grabiner, I. Grattan-Guinness, Jack E. Graver, Daniel Grayson, Richard Green, Curtis Greene, James M. Greenberg, D. S. Griffeath, Richard A. Groeneveld, Benedict H. Gross, Larry C. Grove, Gary Gruenhage, Branko Grünbaum, Victor W. Guillemin, Andrew P. Guinand, Richard Gundy, Ranee Gupta, Karl E. Gustafson, Donald W. Hadwin, James L. Hafner, T. Hailperin, Richard W. Hamming, Bruce Hanson, Denis Hanson, Leon Harkleroad, James Hartman, David Harville, Elmer Hayashi, Eric Hayashi, Maurice Heins, J. M. Henly,

James V. Herod, Gerald Heuer, John C. Higgins, Philip J. Higgins, Theodore P. Hill, Michael Hirsch, Morris W. Hirsch, Ronald Hirshon, Charles R. Hobby, Shu-Ping Hodgson, David Hoff, Michael J. Hoffman, Richard B. Holmes, Thomas B. Hoover, John M. Howie, C. C. Hsiung, John F. Hughes, Rhonda Hughes, Paul Humke, James E. Humphreys, Thomas W. Hungerford, R. Hunt, Joan P. Hutchinson, Edward Ingraham, Kenneth F. Ireland, John R. Isbell, Bradley W. Jackson, James E. Jamison, Gerald T. Janusz, Christian Jensen, Bruce R. Johnson, D. Johnson, Robert H. Johnson, Wells Johnson, Lenny K. Jones, Jean-Pierre Kahone, Michael J. Kallaher, Daw Kalman, David Kammler, Herbert M. Kamowitz, Irving Kaplansky, Edwin H. Kaufman, Robert P. Kaufman, H. Jerome Keisler, Carl Kelley, Leroy M. Kelly, Paul J. Kelly, Kenneth R. Kellum, J. G. Kemeny, John Kenelly, Jeanne Wald Kerr, Clark H. Kimberling, Ronald B. Kirk, David A. Klarner, Richard S. Kleber, Victor L. Klee, John O. Koltinen, Alan G. Konheim, Erwin O. Kreyszig, Richard P. Kubelka, David W. Jeuker, Ray A. Kunze, J. Cleo Kurtz, Richard Laatsch, Carl E. Langenhop, Peter Lappan, Lee M. Larson, Loren C. Larson, David C. Lay, Ronald Leach, Derrick H. Lehmer, Raoul LePage, Fred E. J. Linton, Charles Livingston, Peter A. Loeb, Calvin T. Long, Thomas G. Lucas, Eugene M. Luks, Roger C. Lyndon, Charles MacCluer, Angus J. MacIntyre, Kenneth D. Magill, Jr., Daniel P. Maki, Peter Malcolmson, Joseph Malkevitch, N. S. Mandelsohn, Bennet Manvel, Sandy Markus, Jerrold E. Marsden, William S. Massey, Nelson Max, Jack E. McLaughlin, Daivd G. Mead, Glen Meeden, G. H. Meisters, N. S. Mendelsohn, Russell L. Merris, Ernest A. Michael, George Miel, R. J. Mignone, Dragan Milicic, John A. Mitchem, M. Susan Montgomery, David S. Moore, Ramon E. Moore, Bernard B. Morrel, W. O. J. Moser, James Munkres, Stephen G. Nash, Ho-Guen Ng, Ivan Niven, Katsumi Nomizu, Eric A. Nordgren, Barrett O'Neil, James M. Ortega, Barbara L. Osofsky, John C. Oxtoby, George D. Parker, Beresford N. Parlett, M. Parmenter, Donald S. Passman, Oren Patashnik, E. M. Patterson, S. Das Peddada, Jean Pedersen, Victor C. Pestien, Brian L. Peterson, G. M. Phillips, Veril L. Phillips, Charles Pinter, George Piranian, Jacob M. Plotkin, Harry Pollard, Carl Pomerance, J. B. Pomeranz, Mohsen Pourahmadi, Carl Prather, William M. Priestley, Steven Puckette, Wade Ramey, L. Thomas Ramsey, David F. Rearick, Wulf Rehder, E. M. Reingold, Robert B. Reisel, James Reneke, Bruce A. Reznick, Luis Ribes, J. Ian Richards, Neil W. Richert, Fred Richman, John H. Riley, A. Wayne Roberts, Timothy Robertson, Derek J. S. Robinson, R. Tyrrell Rockafellar, David Roeder, C. Ambrose Rogers, Lajos Ronyai, Andrew Rosalsky, Nicholas J. Rose, David P. Roselle, Michael Rosen, Robert A. Rosenbaum, George Rosenstein, Kenneth A. Ross, John F. Rossi, Gian-Carlo Rota, Joseph J. Rotman, James Rovnyak, Walter Rudin, Sarurnino L. Salas, Donald Sarason, Jonathan Schaer, Juan Jorge Schäffer, Doris J. Schattschneider, Stephen Schecter, Glenn E. Schober, Isaac J. Schoenberg, Bert Schreiber, John Schue, Jacob T. Schwartz, Allen Schweinsberg, Allen J. Schwenk, Jean Seberry, Justus Seeley, George R. Sell, William Shadwick, J. O. Shallit, Joel Shapiro, Burce L. R. Shawyer, Gary J. Sherman, Allen Shields, Rae Michael Shortt, Mohan Shrikande, Neelima Shrikande, Herb Silverman, David A. Singer, Neil J. A. Slane, William R. Sledd, Daniel C. Sloughter, Charles Small, Charles Smith, Philip W. Smith, Wayne Smith, Craig Smorynski, Frederick Solomon, Lee Sonneborn, Terence Paul Speed, Joel H. Spencer, Olaf P. Stackelberg, Joseph Stampfli, Richard St. Andre, Richard P. Stanley, Dennis W. Stanton, Harold W. Stark, Lynn Steen, Robert Steinberg, Kenneth Stephenson, Arthur Harold Stone, David R. Stone, W. Gilbert Strang, Karl R. Stromberg, Walter R. Stromquist, M. V. Subbarao, Carl Sundberg, David Bruce Surowski, Howard S. G. Swann, Laif Swanson, Lajos F. Takács, Richard A. Tapia, Angus E. Taylor, Herbert Taylor, James Thomson, Dmitri E. Thoro, John A. Thorpe, Robert M. Thrall, Lloyd Trefether, Alan Tucker, Thomas W. Tucker, W. T. Tutte, Gerald S. Ungar, Lucien D. van Hamme, Bhushan Lal Wadhwa, Samuel S. Wagstaff, Shelley Walsh, Mitchell Wand, Vladomir A. Waroditsky, Delano Wegener, Clifford E. Weil, Gregory P. Wene, John E. Wetzel, Robert L. Wheeler, Harold Widom, Howard W. Wiener, Thomas W. Wieting, John B. Wilker, Stephen Willard, Hugh C. Williams, Kenneth S. Williams, John Winnie, Warren Wogen, Stephen Wolfram, Edward T. Wong, Alan C. Woods, John W. Wrench, Jr., Steve J. Wright, H. P. Wynn, Frank Young, Lawrence A. Zalcman, Dan Zelinsky, Julius M. Zelmanowitz, Anton Zettl.

Available in Hardcover and Paper Editions



Mathematical Time Exposures

by Issac J. Schoenberg
279 pages
Hardcover Edition: List: $34.50
MAA Member: $26.00
Paper Edition List: $21.00
MAA Member: $15.50

Mathematical Time Exposures was inspired by Hugo Steinhaus' admirable book, *Mathematical Snapshots,* published in 1938. The title, *Mathematical Time Exposures* was also suggested by photography, but Schoenberg's pace is much more leisurely than Steinhaus'. Schoenberg spends more time on fewer subjects—the "snapshots" become "time exposures." The subject of at least two of the chapters actually antedate the invention of the daguerreotype.

The author manages to bring together concepts from geometry, number theory, algebra, and analysis, frequently mixing them together in the same chapter. The arts are not neglected. Discussions on the tuning of keyboard instruments, the guitar, and the vibrations of strings are discussed, as well as the suggestion of rectilinear models for outdoor sculpture.

Many of the chapters in this book may be read independently, and at least half of them with only a knowledge of precalculus mathematics.

THE MATHEMATICAL ASSOCIATION OF AMERICA

ORDER FROM:
The Mathematical Association of America
1529 Eioghteenth St, NW
Washington, DC 20036

HM

College Algebra
Timothy J. Kelly, Hamilton College
Richard H. Balomenos
University of New Hampshire
John T. Anderson
Hamilton College

About 464 pages • cloth • Study Guide • Instructor's Annotated Edition • Instructor's Alternate Testing Program • Solutions Manual • Computerized Test Generator • Early 1986

College Algebra features a goal-referenced organization that ties instruction, practice, and evaluation into a cohesive whole. Other noteworthy features are annotated worked examples, clear exposition, and abundant exercises, with an emphasis on applications throughout. All definitions are mathematically complete, correct, and concise. Students will find the problems imaginative and challenging.

Intermediate Algebra and Trigonometry with Applications Revised Edition
M. Wiles Keller, Purdue University

About 640 pages • paper
Instructor's Annotated Edition
Alternate Testing Program
Early 1986

Keller provides—in a single text—a review of both second-year algebra and trigonometry. His step-by-step technique features an annotated worked example of each skill and application. An objective-referenced organization links lessons, practice, and testing. Throughout the text an emphasis on applications lets students use acquired skills in practical contexts. Included are nearly 5,000 problems.

Basic Mathematics and Algebra with Applications Fourth Edition
M. Wiles Keller, Purdue University
James H. Zant
Oklahoma State University

546 perforated pages • paper
Instructor's Annotated Edition
Alternate Testing Program • 1984

Keller and Zant's developmental mathematics program reviews both basic mathematics and elementary algebra within a single text. Step-by-step instruction and objective-referenced organization —as described for the Keller text above—helps students help themselves. An abundance of problems in situational settings provides the student with opportunities to use skills as they are learned.

Handbook of Basic Math with Applications
James H. Zant and **M. Wiles Keller**

198 perforated pages • paper
Answer Booklet • 1985

An economical review of the basic math portion of the above text.


Here are the solutions to
your textbook problems...
Discrete Mathematics
with Computer Science
Applications
SKVARCIUS / ROBINSON
Fundamentals
of
Differential
Equations
R. KENT NAGLE
EDWARD B. SAFF
MATHEMATICS

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, DC 20036

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 93, Number 4 April 1986

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. (This is necessary for indexing purposes.) Send all main articles, letters to the editor, and editorial correspondence to the editor-elect: H.S. WILF, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Contributions to the other four special departments (Notes, Teaching of Mathematics, Unsolved Problems, and Problems and Solutions) should be sent to the appropriate one of the associate editors, as follows:

Notes: CAROL G. CRAWFORD, Department of Mathematics, U.S. Naval Academy, Annapolis, MD 21402, *or* RICHARD LIBERA, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716, *or* ANITA E. SOLOW, Department of Mathematics, Grinnell College, Grinnell, IA 50112.

Teaching of Mathematics: JOAN P. HUTCHINSON *or* STAN WAGON, Department of Mathematics, Smith College, Northampton, MA 01063.

Unsolved Problems: RICHARD GUY, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.

Proposed problems and solutions, both elementary and advanced: PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801.

Reprint Permission: A. B. WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Back Issues: P. and H. BLISS Co., Middletown, CT 06457.

Microfilm Editions: University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

Editorial Assistants: Fanny Hailpern, Mary P. Jackson.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT, ten times a year: January, February, March, April, May, June-July, August-September, October, November and December.

The annual subscription price for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is $24 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $24. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $70 per year.

Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

PRINTED IN THE UNITED STATES OF AMERICA

AWARD FOR DISTINGUISHED SERVICE TO PROFESSOR ARNOLD EPHRAIM ROSS

ANNELI LAX
Department of Mathematics, NYU-Courant Institute, New York, NY 10012

ALAN C. WOODS
Department of Mathematics, Ohio State University, Columbus, OH 43210

The 1985 recipient of MAA's Award for Distinguished Service to Mathematics is Arnold Ephraim Ross who has devoted his life to Science Education through his unique summer program for high school students. He has profoundly influenced many people early in their lives, among them a great number of our most original, now eminent, colleagues in mathematics. He receives this award for his significant impact on mathematics, via mathematics education, on a national scale. Indeed no major mathematics conference is without a knot of mathematicians who compare notes on their experiences in "The Summer Program".

Arnold E. Ross was born in Chicago in 1906. In his youth his family moved to Russia. He completed his precollege education there and was admitted to the University of Odessa at age 16. He studied mathematics and met Felix Gantmacher, one of three other teen-age students attending lectures at Odessa. After returning to the U.S.A. Ross received his B.S., M.S. and Ph.D. degrees in mathematics from the University of Chicago. There he came under the influence of E. H. Moore and his problem-oriented "discovery" method of teaching, which Ross later carried into his own teaching. Ross wrote his dissertation in 1931 under L. E. Dickson.

This was not a good year for getting into any job market, yet Ross obtained a then rare National Research Council fellowship and spent two years at CalTech. He returned to Chicago to head, from 1933 to 1935, the Mathematics Department of People's Junior College, an experimental cooperative venture by some young Ph.D.'s during the Depression. He subsequently joined the faculty of St. Louis University until 1946, when he became Professor and Chair of the Mathematics Department at Notre Dame. In 1963 he was recruited to reorganize the Department of Mathematics at the Ohio State University where he remained as chairman until he retired and was named Professor Emeritus in 1976.

His remarkable educational activities had their formal beginning in 1957 when he responded to the post-Sputnik shortage of mathematics and science teachers by creating a graduate teachers' training program at Notre Dame. He continued this summer program at Notre Dame for the next six years, gradually changing the emphasis from the training of high school teachers to the training of high school students, and then transplanting it to Ohio State where it flourished. In 1975 the University of Chicago hosted the program, with Ross as director. It stayed in Chicago until 1979 then returned again to the Ohio State campus, and there it remains to date.

In its initial Notre Dame period, Ross's program featured courses in number theory, classical analysis, linear algebra and geometry. But Ross soon recognized elementary number theory as the vehicle most suited for gifted, but inexperienced high school students. There are morning lectures on number theory, afternoon problem seminars, and lots of carefully structured problem sets geared to exploration and discovery.

A crucial part in Ross's program is played by former participants who return to counsel and serve as role models to younger ones, and who are offered elective courses in more advanced topics.

In addition to providing these intense mathematical summer experiences, Ross has created several other programs; for example, he initiated Horizons Unlimited in 1970, designed to uncover and develop talents of inner city school children from 6th to 12th grade in Columbus, Ohio. He served as consultant to a program in India for gifted children during the Spring of 1973. In 1975

ARNOLD EPHRAIM ROSS

he accepted an invitation by the Australian National University to take part in a summer talent search modeled on his Ohio State Program and remained active in that January summer program until 1983. Between summer programs in the northern and southern hemispheres, he found time in 1978 and 1979 to help initiate, with Professor P. Roquette, a similar program in Heidelberg.

Arnold Ross has been widely recognized as an outstanding mathematics educator. The long list of his achievements, service and awarded honors include membership of NSF's Advisory Board on Science Education and of CUPM where he chaired a panel on innovations, his being named Outstanding Educator in America (1974–75), receiving Notre Dame's Award of Honor on its 1965 Centennial of Science, Ohio State's Distinguished Teaching and Service Awards, and an Honorary Doctorate from Denison University in 1984.

But the greatest tribute to him is the respect, affection and recognition given him by his former students and by some of our colleagues who were fortunate enough to teach his former students when they reached college or graduate school.

The following small sample of their comments is particularly refreshing for two deep reasons: it gives testimony to the long range effects of an educational experience on people's lives rather than on their end-of-year test scores; and it includes people in many professions, not just mathematicians.

We learned that the fun of doing mathematics—and, by extension, of engaging in other intellectual activities as well—is directly proportional to the intensity of one's commitment.

* * *

Some of us pursue careers in mathematics; others have chosen divergent paths. But there can be few indeed who participated in Dr. Ross's summer programs without being profoundly affected by the experience.

* * *

A. Ross's summer program has no equals in terms of inspiring students with the spirit of mathematical research and individual creativity. He has set an example few of us can meet, but which is of decisive importance to the future and ethos of American Mathematical Education.

HARMONIC FUNCTIONS FROM A COMPLEX ANALYSIS VIEWPOINT

SHELDON AXLER

Department of Mathematics, Michigan State University, East Lansing, MI 48824

Recall that a real valued function u defined on an open set in the complex plane is called harmonic if the partial second derivatives of u exist and are continuous and

$$u_{xx} + u_{yy}$$

is identically zero.

Although harmonic functions play a crucial role in several areas of pure and applied mathematics, most mathematicians and mathematics students are more familiar with the elementary properties of analytic functions than of harmonic functions. For example, almost all of us have had to take a test some time in our lives in which we needed to know that an isolated

Sheldon Axler received an A.B. degree from Princeton University in 1971 and a Ph.D. in mathematics in 1975 from the University of California, Berkeley, where he was a student of Donald Sarason. He has taught at M.I.T., Indiana University, and Michigan State University. His research centers around the interplay between complex analysis and operator theory. His nonmathematical interests include travel, running, history, and world affairs.

singularity of a bounded analytic function is removable. As we will see later, the same theorem holds for harmonic functions, yet this result is not common knowledge among nonspecialists. This paper will study harmonic functions from the comfortable perspective of complex analysis.

We all know that the real and imaginary parts of an analytic function are harmonic. On an open disk the converse is true—every real valued harmonic function is the real part of some analytic function. With this knowledge, the reader should be able to derive most of the *local* properties of harmonic functions from the corresponding properties of analytic functions. For example, every real valued harmonic function is infinitely differentiable, satisfies the mean value property, cannot take on a local maximum or minimum without being constant, etc.

The global behavior of harmonic functions is not so simple to analyze, because on an arbitrary region not every harmonic function is the real part of some analytic function. For example, consider the function $\log|z|$ defined on some open annulus centered at the origin. This function is harmonic, but it is not the real part of any function analytic on the annulus. As we will see later, this is essentially the only example of this phenomenon on an annulus, in the sense that any real valued harmonic function on an annulus is the real part of some function analytic on the annulus plus a real constant times $\log|z|$.

The main theme of this paper is that on finitely connected regions every real valued harmonic function is the real part of some function analytic on the region plus some logarithm terms; the Logarithmic Conjugation Theorem in the next section will give a precise statement. For our purposes, it is best to define a finitely connected region Ω to be a nonempty, connected open subset of the complex plane whose complement (with respect to the complex plane) has only finitely many bounded connected components. For obvious geometrical reasons (see Fig. 1), each bounded component of the complement of Ω is called a hole of Ω. A finitely connected region that has no holes is of course called simply connected.

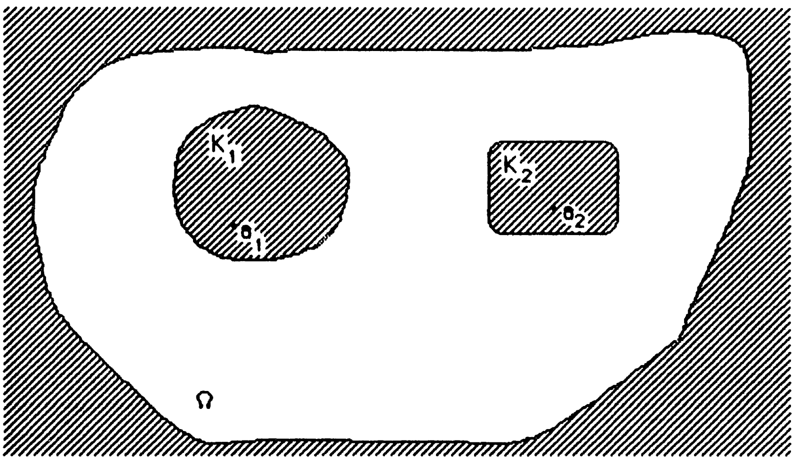


FIG. 1

Harmonic Conjugates. If Ω is simply connected and u is a real valued harmonic function on Ω, then there is an analytic function f on Ω such that $u = \operatorname{Re} f$; here Re denotes the real part, and Im will denote the imaginary part. The function $\operatorname{Im} f$ is called the harmonic conjugate of u, and except for the addition of a constant, it is unique.

Harmonic conjugates are so useful that many mathematicians use them even when they don't exist (on regions that are not simply connected). Thus it is possible to read about a harmonic conjugate that is a "multi-valued function with periods." I have always been mystified by this terminology. Any good high school student knows that if g is a function and z is an element of the domain of g, then $g(z)$ is a single object. Thus, by definition, a function cannot be

multi-valued. Complex analysts will recognize that Riemann surfaces provide the rigorous mathematics lurking behind the "multi-valued functions" that appear. However, "multi-valued functions" are often discussed in texts where the reader is unfamiliar with the concept of a Riemann surface, and I suspect that many people (including me) are confused by the nebulous concept of a "multi-valued function." The approach I will present here allows us to obtain the results we want without resorting to either "multi-valued functions" or Riemann surfaces.

The following theorem, which is the main tool of this paper, can replace the "multi-valued function" approach to harmonic conjugation and simplify many proofs. Since this theorem states that each function on a finitely connected region has a harmonic conjugate, provided we first subtract some logarithmic terms, I call it the Logarithmic Conjugation Theorem.

LOGARITHMIC CONJUGATION THEOREM. *Suppose Ω is a finitely connected region, with $K_1, \dots, K_N$ denoting the bounded components of the complement of Ω. For each j, let a_j be a point in K_j. If u is a real valued harmonic function on Ω, then there exist an analytic function f on Ω and real numbers $c_1, \dots, c_N$ such that*

$$u(z) = \operatorname{Re} f(z) + c_1 \log|z - a_1| + \cdots + c_N \log|z - a_N|$$

for every z in Ω.

With this theorem, there is no need to talk about "the periods of the multi-valued harmonic conjugate of u," but if one were to do so, the periods would turn out to be exactly the numbers $c_1, \dots, c_N$.

The only place I have been able to find the Logarithmic Conjugation Theorem written down with a proof is in an old paper of Walsh [**18**], pages 518 and 527. Walsh's proof uses a special version of Green's formula that is valid for harmonic functions. The elementary proof I will present here uses only the Cauchy Integral Theorem for analytic functions, so it should be more accessible to modern audiences. After proving the Logarithmic Conjugation Theorem, we will see how it can be used to study isolated singularities of harmonic functions, help solve the Dirichlet problem on annuli, and lead to a short proof of the conformal mapping theorem for doubly connected regions.

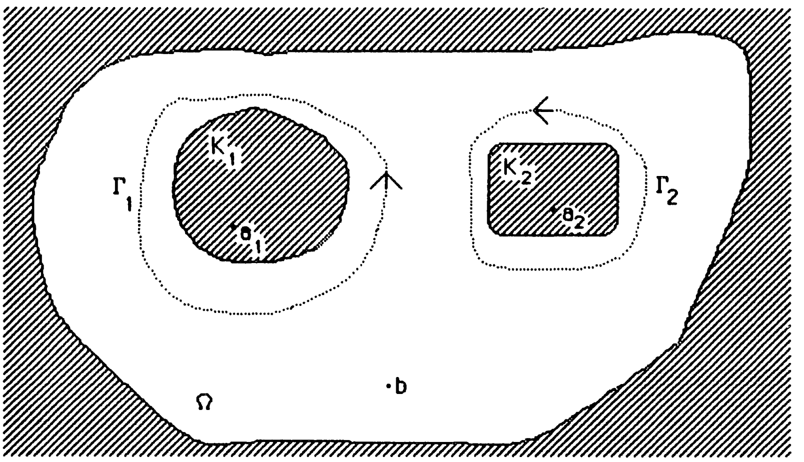


FIG. 2

Proof of the Logarithmic Conjugation Theorem. Define a function h on Ω by

$$h(z) = u_x(z) - iu_y(z).$$

The function h is analytic on Ω; this is proved by verifying the Cauchy-Riemann equations, which in the case of h are

$$u_{xx} = -u_{yy},$$
$$u_{xy} = -(-u_{yx}).$$

Of course, the first equation holds because u is harmonic, and the second equation holds because the order of partial differentiation doesn't matter.

For each j $(1 \leqslant j \leqslant N)$, let Γ_j be a curve in Ω that surrounds the hole K_j; see Fig. 2. (The technical definition of "surrounds" is that Γ_j has winding number one about each point of K_j and winding number zero about all the other holes of Ω.) Now define c_j by

$$c_j = (1/2\pi i)\int_{\Gamma_j} h(w)\, dw.$$

To see that each c_j is a real number, note that

$$\begin{aligned}\operatorname{Im} c_j &= (-1/2\pi)\operatorname{Re}\int_{\Gamma_j} h(z)\, dz \\ &= (-1/2\pi)\operatorname{Re}\int_{\Gamma_j} (u_x - iu_y)(dx + idy) \\ &= (-1/2\pi)\int_{\Gamma_j} u_x\, dx + u_y\, dy.\end{aligned}$$

The last integral is an exact differential over a closed curve, so it equals zero, and thus c_j is real.

Now fix a point b in Ω, and define a function f on Ω by

$$f(z) = \int_b^z h(w) - \frac{c_1}{w - a_1} - \cdots - \frac{c_N}{w - a_N}\, dw,$$

where the integration is taken over any path in Ω from b to z. To show that f is well defined, we must check that the integral above is independent of the path from b to z. But given two different paths from b to z, we can reverse the direction of the second path, getting a closed curve in Ω; see Fig. 3.

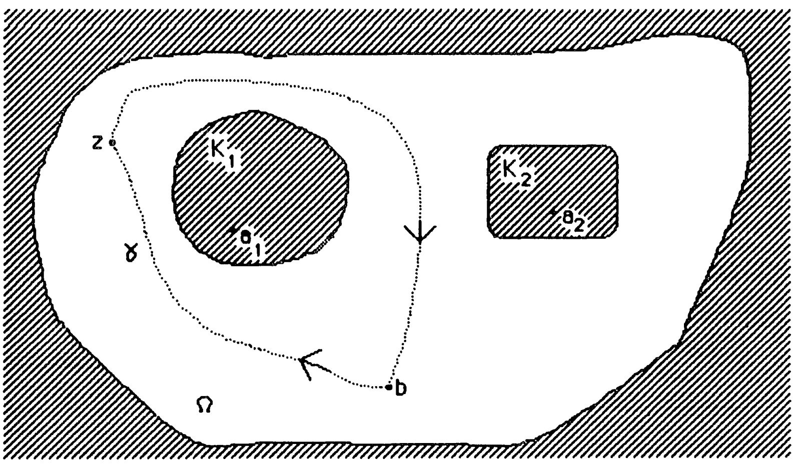


FIG. 3

Thus we need only show that

$$(*) \qquad 0 = \frac{1}{2\pi i}\int_\gamma h(w) - \frac{c_1}{w - a_1} - \cdots - \frac{c_N}{w - a_N}\, dw$$

for each closed curve γ in Ω. However, if m_j denotes the winding number of γ about K_j, then the Cauchy Integral Theorem (along with the definition of c_j) says that

$$(1/2\pi i)\int_\gamma h(w)\,dw = m_1c_1 + \cdots + m_Nc_N,$$

while the definition of winding number implies that

$$\frac{1}{2\pi i}\int_\gamma \frac{c_1}{w-a_1} + \cdots + \frac{c_N}{w-a_N}\,dw = m_1c_1 + \cdots + m_Nc_N.$$

The last two equations show that $(*)$ holds.

Now that we know f is well defined, it is clear (just examine the difference quotient) that f is analytic on Ω and

$$f'(z) = h(z) - c_1/(z-a_1) - \cdots - c_N/(z-a_N).$$

Having assembled our analytic function f and real numbers $c_1, \dots, c_N$, we are ready to verify that the conclusion of the Logarithmic Conjugation Theorem holds. For convenience, define a function q on Ω by

$$q(z) = \operatorname{Re} f(z) + c_1\log|z-a_1| + \cdots + c_N\log|z-a_N|.$$

Add a constant to f so that u and q agree at some point of Ω, say at b. We will finish the proof by showing that $u_x(z) = q_x(z)$ and $u_y(z) = q_y(z)$ for all points z in Ω, so that $u = q$.

Differentiation is a local operation, and since locally $\log(z-a_j)$ is analytic on Ω, we have

$$\begin{aligned} q_x(z) &= \left[\operatorname{Re} f(z) + c_1\log|z-a_1| + \cdots + c_N\log|z-a_N|\right]_x \\ &= \left[\operatorname{Re}\{f(z) + c_1\log(z-a_1) + \cdots + c_N\log(z-a_N)\}\right]_x \\ &= \operatorname{Re}\left[\{f(z) + c_1\log(z-a_1) + \cdots + c_N\log(z-a_N)\}_x\right]. \end{aligned}$$

For analytic functions, the partial derivative with respect to x is the same as the complex derivative, so the last equation becomes

$$q_x(z) = \operatorname{Re}\left[f'(z) + c_1/(z-a_1) + \cdots + c_n/(z-a_n)\right].$$

Plugging the formula we found earlier for f' into the above equation gives

$$q_x(z) = \operatorname{Re} h(z),$$

and recalling the definition of h, we get

$$q_x(z) = u_x(z).$$

Finally, computing q_y is done in the same fashion, except now we note that taking the y partial derivative of an analytic function gives i times the complex derivative. So we get

$$q_y(z) = \operatorname{Re}(ih(z)),$$

and again referring to the definition of h gives

$$q_y(z) = u_y(z).$$

Thus we have completed the proof of the Logarithmic Conjugation Theorem.

Isolated Singularities. To study the behavior of harmonic functions near isolated singularities, we need only consider harmonic functions defined on punctured disks. So let $\mathbb{D}$ denote the open unit disk in the complex plane, and let $\mathbb{D}'$ denote the open unit disk with the origin removed. The complement of $\mathbb{D}'$ has only one bounded component. Thus the Logarithmic Conjugation Theorem (with $a_1 = 0$) tells us that any real valued harmonic function u on $\mathbb{D}'$ can be written in the form

$$u(z) = \operatorname{Re} f(z) + c \log |z|,$$

where f is analytic on $\mathbb{D}'$ and c is a real number. Using the classification of isolated singularities of analytic functions into removable singularities, poles, and essential singularities, we can now see the possible behaviors of harmonic functions near isolated singularities.

In particular, suppose our harmonic function u is bounded on $\mathbb{D}'$. Rewrite the above equation as

$$\operatorname{Re} f(z) = u(z) - c \log |z|.$$

Suppose first that c is positive. Since u is bounded, this means that $\operatorname{Re} f(z) \to \infty$ as $z \to 0$. This behavior is clearly impossible if f has a removable singularity at 0. Also, f cannot have a pole at 0, because then $f(\mathbb{D}')$ would include the complement of some disk, and in particular $f(\mathbb{D}')$ would include a sequence whose real part tends to $-\infty$. Finally, f cannot have an essential singularity at 0, because then there would be a sequence tending to 0 on which f (and hence $\operatorname{Re} f$) stays bounded. Since no other possibilities remain, we conclude that c cannot be positive. Similarly, c cannot be negative. Thus $c = 0$ and so $u = \operatorname{Re} f$. This means that $\operatorname{Re} f$ is bounded. Again this is impossible if f has either a pole or an essential singularity at 0, and so f has a removable singularity at 0. Of course this means that u also has a removable singularity at 0. We have thus proved the following theorem.

ISOLATED SINGULARITY, BOUNDED FUNCTION. *If a harmonic function is bounded near an isolated singularity, then that singularity is removable.*

The above theorem was first published by Schwarz [**16**], p. 252, whose proof used the series representation for harmonic functions on annuli (discussed in the next section of this paper). Many years later Picard [**13**], who seemed not to know that Schwarz had already published a proof, gave a proof using a "multi-valued harmonic conjugate with a period," precisely the object that the Logarithmic Conjugation Theorem allows us to avoid. Perhaps Lebesgue disliked "multi-valued functions"; three weeks after Picard presented his proof to the Académie des Sciences in Paris, Lebesgue (also unaware of Schwarz's proof) presented another proof [**11**] to the same body. Lebesgue's proof requires knowledge of the solvability of the Dirichlet problem for an annulus (discussed in the next section of this paper).

Now let's see how the Logarithmic Conjugation Theorem can help in the study of some important spaces of analytic functions. For an arbitrary region Ω and a positive number p, the much studied Hardy space $H^p(\Omega)$ is defined to be the set of analytic functions h on Ω such that there exists a harmonic function u on Ω with

$$|h(z)|^p \leqslant u(z)$$

for all z in Ω. If Ω is the unit disk $\mathbb{D}$, this definition is equivalent to the usual definition involving integrals around concentric circles. A good reference on Hardy spaces is Fisher's recent book [**6**].

The Logarithmic Conjugation Theorem can be especially useful when investigating properties of Hardy spaces on finitely connected regions. Here is an example of an important theorem from that subject: Let Ω be a bounded finitely connected region with smooth boundary. Then each continuous real valued function on $\partial\Omega$ can be uniformly approximated by functions of the form

$$\operatorname{Re} f(z) + c_1 \log |z - a_1| + \cdots + c_N \log |z - a_N|,$$

where f is a rational function whose poles are all in the exterior of Ω, where each c_j is a real number, and where each a_j is in the interior of the corresponding hole of Ω (see Fig. 1).

A proof of the result above (see for example [**6**], Theorem 2.1) would take us beyond the scope of this paper, but the reader who has noted that the theorem above is formally similar to the Logarithmic Conjugation Theorem will not be surprised that the Logarithmic Conjugation Theorem is useful in simplifying the standard proof.

A function in a Hardy space need not be bounded, so the following theorem is stronger than the usual result about isolated singularities of bounded analytic functions. As we will see in the proof, the Logarithmic Conjugation Theorem is an excellent tool for dealing with isolated singularities of Hardy space functions.

ISOLATED SINGULARITY; HARDY SPACE FUNCTION. *An isolated singularity of a function that belongs to a Hardy space is removable.*

To prove this theorem, we can fix a positive number p and assume that h is a function in $H^p(\mathbb{D}')$; we need to show that h is analytic at 0. By definition, there is a function u harmonic on $\mathbb{D}'$ such that

$$|h(z)|^p \leqslant u(z)$$

for each z in $\mathbb{D}'$. As usual, use the Logarithmic Conjugation Theorem to write u in the form

$$u(z) = \operatorname{Re} f(z) + c \log |z|,$$

where f is analytic on $\mathbb{D}'$ and c is a real number. Now

$$|e^{-f(z)}| = e^{-\operatorname{Re} f(z)} = e^{-u(z) + c \log |z|} = |z|^c e^{-u(z)},$$

so

$$|e^{-f(z)}|\,|z|^{-c} = e^{-u(z)} \leqslant 1.$$

If f had either a pole or an essential singularity at 0, then there would be a sequence tending to 0 on which e^{-f} is bounded and also a sequence tending to 0 on which e^{-f} is unbounded; this behavior would mean that e^{-f} has an essential singularity at 0. However, the last line of the previous paragraph shows that e^{-f} has either a pole or a removable singularity at 0. Thus we can conclude that f has a removable singularity at 0.

So now we know that

$$|h(z)|^p \leqslant \operatorname{Re} f(z) + c \log |z|,$$

where f is analytic on $\mathbb{D}$. Since f is bounded near 0, the above inequality implies that $|z|^p |h(z)|^p \to 0$ as $z \to 0$. Thus $zh(z) \to 0$ as $z \to 0$, which implies that h is analytic at 0, as desired.

Analysis on Annuli. From now on, $\mathscr{A}$ will denote an open annulus centered at the origin. In this section, we will use the Logarithmic Conjugation Theorem to investigate the integrals of a harmonic function around concentric circles and to give a series representation for the most general harmonic function on $\mathscr{A}$. Then we will use this information to guess the solution to the Dirichlet problem for annuli. Until we begin talking about the Dirichlet problem, everything that is said about annuli will also apply to punctured disks.

If u is a real valued harmonic function on $\mathscr{A}$, then the Logarithmic Conjugation Theorem tells us that there is a function f analytic on $\mathscr{A}$ and a real number c such that

$$u(z) = \operatorname{Re} f(z) + c \log |z|$$

for every z in $\mathscr{A}$.

Suppose we fix a positive number r in $\mathscr{A}$. Then from the equation above we get

$$\begin{aligned}(1/2\pi)\int_0^{2\pi} u(re^{i\theta})\, d\theta &= c \log r + \operatorname{Re}\left[(1/2\pi)\int_0^{2\pi} f(re^{i\theta})\, d\theta\right] \\ &= c \log r + \operatorname{Re}\left[(1/2\pi i)\int_{\partial(r\mathbb{D})} f(z)/z\, dz\right],\end{aligned}$$

where as usual $\mathbb{D}$ denotes the unit disk. The Cauchy Integral Theorem tells us that the last term in the equation above is independent of r (in fact, the last term equals the real part of the constant

term in the Laurent series expansion of f). Thus we have the following result:

INTEGRALS AROUND CIRCLES. *If u is harmonic on $\mathscr{A}$, then*

$$\int_0^{2\pi} u(re^{i\theta})\,d\theta$$

is a linear function of $\log r$.

The fact above has been noted elsewhere (see [**1**], Theorem 20, p. 265 and [**15**], p. 21), but the proof given here using the Logarithmic Conjugation Theorem seems considerably easier than other proofs.

Now let's use the Logarithmic Conjugation Theorem to find a series representation for the most general real valued harmonic function u on $\mathscr{A}$. As usual, first we write

$$u(z) = \operatorname{Re} f(z) + c \log|z|.$$

Since f is analytic on the annulus $\mathscr{A}$, we can express f as a Laurent series:

$$f(z) = \sum_{n=-\infty}^{\infty} 2a_n z^n,$$

where for convenience the coefficient of z^n is called $2a_n$. Now

$$u(z) = \left[f(z) + \overline{f(z)}\right]/2 + c\log|z|.$$

Let $z = re^{i\theta}$, replace $f(z)$ with its Laurent series, and evaluate $\overline{f(z)}$ by taking the complex conjugate of the Laurent series for $f(z)$, and then replace n with $-n$ in the summation, getting the following formula:

SERIES REPRESENTATION. *If u is a real valued harmonic function on an annulus $\mathscr{A}$, then u has the form*

$$(**)\qquad u(re^{i\theta}) = c\log r + \sum_{n=-\infty}^{\infty}\left(a_n r^n + \overline{a_{-n}}\, r^{-n}\right)e^{in\theta}.$$

The infinite sum above converges absolutely for each $re^{i\theta}$ in $\mathscr{A}$ and uniformly on compact subsets of $\mathscr{A}$ (because the Laurent series for f has these properties). Other derivations of this useful formula are given by Saks and Zygmund [**14**], Chapter 10, Section 3, and Heins [**8**], pages 56–57.

The above formula can help us see that the Dirichlet problem is solvable for annuli. Recall that for a bounded region Ω, we say the Dirichlet problem is solvable on Ω if for each continuous real valued function U on $\partial\Omega$, there is a continuous real valued function u on the closure of Ω such that u is harmonic on Ω and $u|\partial\Omega = U$.

For the unit disk $\mathbb{D}$, the Dirichlet problem is explicitly solved by the Poisson integral expressing u in terms of U:

$$u(re^{i\theta}) = \frac{1-r^2}{2\pi}\int_0^{2\pi}\frac{U(e^{i(\theta-t)})}{1-2r\cos t + r^2}\,dt.$$

The example of the disk leads us to suspect that for an annulus $\mathscr{A}$, we could find u from U by integrating U on the two circles of $\partial\mathscr{A}$ against an appropriate kernel. This is correct, but the kernel for $\mathscr{A}$ is not an elementary function (like $(1-r^2)/[1-2r\cos t + r^2]$ for the disk), but a more complicated expression involving elliptic functions; see [**17**] and [**10**]. Alternatively, the annulus can be lifted to its universal covering space (represented as an infinite strip) and then we can integrate on lines rather than circles. This procedure, worked out by Sarason [**15**], pages 21–23, and more recently by Wang [**19**], Theorem 1, gives a kernel on the boundary of the strip that is an elementary (but complicated) function.

We will take a different approach here, using the Series Representation to lead us to an

elementary proof that the Dirichlet problem is solvable for an annulus. To begin, let s denote the inner radius of $\mathscr{A}$ and t denote the outer radius of $\mathscr{A}$. Consider a real valued function U on $\partial\mathscr{A}$ of the form

$$U(se^{i\theta}) = \sum_{n=-N}^{N} b_n e^{in\theta}$$

$$U(te^{i\theta}) = \sum_{n=-N}^{N} d_n e^{in\theta}.$$

Since U is real valued, b_{-n} must equal the complex conjugate of b_n (just equate U to $\overline{U}$), and similarly for the coefficients d_n. We will first show that the Dirichlet problem is solvable for functions U of the above form, and then show why this implies that the Dirichlet problem is solvable for every continuous real valued function on $\partial\mathscr{A}$.

Compare our formula $(**)$ for the most general harmonic function on $\mathscr{A}$ with the expressions defining U; it is clear that for n positive we want to choose a_n and a_{-n} so that

$$a_n s^n + \overline{a_{-n}}\, s^{-n} = b_n,$$

$$a_n t^n + \overline{a_{-n}}\, t^{-n} = d_n.$$

For each positive n, we can consider these equations as two linear equations (with a_n and $\overline{a_{-n}}$ being the unknown quantities). Since $s \neq t$, the Jacobian of this system is nonzero, and so we can solve for a_n and $\overline{a_{-n}}$.

The relationship between b_n and b_{-n} (and d_n and d_{-n}) insures that we would get the same values for a_n and a_{-n} if we considered these equations for n negative. What happens when n equals zero? Then the Jacobian of the system is 0, and there is not necessarily a solution to the equations. However, we are rescued by the logarithm term in the formula $(**)$ for a harmonic function on $\mathscr{A}$. Now the equations we need to solve are

$$c \log s + 2a_0 = b_0,$$

$$c \log t + 2a_0 = d_0.$$

These two equations (with c and a_0 being the unknown quantities) have a nonzero Jacobian, and so we can solve for c and a_0.

At this point we know $\{a_n\}$ and c; the function u that solves the Dirichlet problem for U is just

$$u(re^{i\theta}) = c \log r + \sum_{n=-N}^{N} \left(a_n r^n + \overline{a_{-n}}\, r^{-n}\right) e^{in\theta}.$$

The interested reader can simply solve the equations to write $\{a_n\}$ and c explicitly in terms of $\{b_n\}$ and $\{d_n\}$.

Now that we know the Dirichlet problem is solvable for functions of the form considered above, we can easily deal with an arbitrary continuous real valued function U on $\partial\mathscr{A}$. By the Stone-Weierstrass Theorem, there is a sequence $\{U_n\}$ of functions on $\partial\mathscr{A}$ of the above form such that U_n converges uniformly to U on $\partial\mathscr{A}$. Let u_n denote the solution to the Dirichlet problem for U_n. The maximum modulus theorem for harmonic functions implies that $\{u_n\}$ is a uniform Cauchy sequence on the closure of $\mathscr{A}$, and it is clear that the limit of this sequence solves the Dirichlet problem for U, and so we are done.

Heins [**9**], pages 301–302, also uses the Stone-Weierstrass Theorem to pass from trigonometric polynomials to arbitrary functions in solving the Dirichlet problem. The advantage of the approach presented here is that the Series Representation allows us to guess naturally the solution for the trigonometric polynomials.

Rather than show that the Dirichlet problem is solvable on $\mathscr{A}$ by dealing with a dense set of

functions, an alternative would be to solve directly the problem for an arbitrary continuous function on $\partial\mathscr{A}$, although the verification requires some knowledge about Fourier series. The idea, still motivated by the Series Representation, is to take an arbitrary continuous function U on $\partial\mathscr{A}$, and write it in the form

$$U(se^{i\theta}) = \sum_{n=-\infty}^{\infty} b_n e^{in\theta},$$

$$U(te^{i\theta}) = \sum_{n=-\infty}^{\infty} d_n e^{in\theta}.$$

Here the infinite sum converges in the L^2 norm (with respect to arc length measure on $\partial\mathscr{A}$). Now simply solve the same equations as with the previous method to find $\{a_n\}$ and c. The harmonic function u on $\mathscr{A}$ that solves the Dirichlet problem for U is then given by

$$u(re^{i\theta}) = c\log r + \sum_{n=-\infty}^{\infty} \left(a_n r^n + \overline{a_{-n}} r^{-n}\right) e^{in\theta}.$$

Conformal Mapping. A region Ω is called doubly connected if its complement (with respect to the complex plane) has precisely one bounded connected component. Of course every annulus is doubly connected. Two annuli are conformally equivalent if and only if their outer radius/inner radius ratios are equal (see [7], Chapter 5, Section 1, Theorem 2).

In this section we will use the Logarithmic Conjugation Theorem to give an easy proof that any doubly connected region is conformally equivalent to some annulus. So that we can have a clean statement of the correct theorem, for this section an annulus is defined to be a set of the form $\{z \in \mathbb{C}: s < |z| < t\}$, where s is a non-negative number (possibly 0) and t is a positive number or ∞.

DOUBLY CONNECTED MAPPING THEOREM. *Let Ω be a doubly connected region. Then there is an annulus $\mathscr{A}$ and a one-to-one analytic function g on Ω that maps Ω onto $\mathscr{A}$.*

The proof of the Doubly Connected Mapping Theorem given by Courant [5], Chapter 1, Section 7, is somewhat similar in spirit to the proof presented here, except that Courant uses a "multi-valued harmonic conjugate with a period," which the Logarithmic Conjugation Theorem will allow us to avoid. Bieberbach [2], Chapter 5, Section 26, gives a proof using Riemann surfaces (and leaves out the hard details). Other proofs of the Doubly Connected Mapping Theorem can be found in the books of Nehari [12], Chapter 7, and Goluzin [7], Chapter 5, Section 1.

In the last section we showed that the Dirichlet problem is solvable on nondegenerate annuli. For our proof of the Doubly Connected Mapping Theorem we will need to know that the Dirichlet problem is solvable for any bounded doubly connected region with smooth boundary. This is really not so difficult to prove. Proofs that fit within the context of a standard complex variables course are given in the texts of Ahlfors [1], Chapter 6, Theorem 9, and Conway [4], Chapter 10, Corollary 4.17. Browder [3], Theorem 3.4.10, gives a proof that is easier, although it requires more machinery.

Proof of the Doubly Connected Mapping Theorem. As usual, we can assume that our doubly connected region Ω is bounded and has a smooth boundary. The standard easy proof that any doubly connected region, except the punctured disk or the punctured plane, is conformally equivalent to a bounded region with smooth boundary can be found in [1], pages 252–253. Actually, our proof does not really require a smooth boundary; however, this assumption does make it easier to prove that the Dirichlet problem is solvable for Ω.

Let K denote the bounded component of $\mathbb{C} \sim \Omega$. Without loss of generality we can assume that $0 \in K$. For obvious reasons, we refer to ∂K as the inner boundary of Ω and $\partial(\Omega \cup K)$ as the outer boundary of Ω; see Fig. 4. Let u be the continuous real valued function on the closure

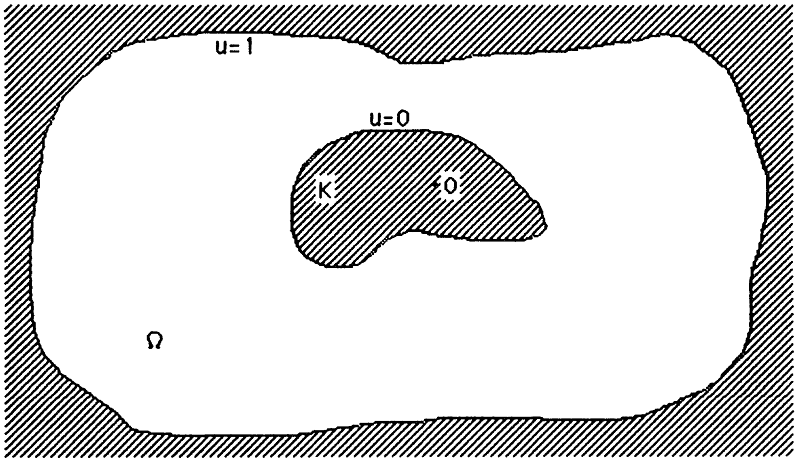


FIG. 4

of Ω that equals 0 on the inner boundary of Ω and equals 1 on the outer boundary of Ω and is harmonic on Ω. Use the Logarithmic Conjugation Theorem to write u in the form

$$u(z) = c \log |z| + \operatorname{Re} f(z),$$

where c is a real number and f is analytic on Ω.

Define a function g on Ω by

$$g(z) = ze^{f(z)/c}.$$

Of course g is analytic on Ω (because f is analytic), and we will see that g maps Ω onto an annulus in a one-to-one fashion. (Let's assume that c is positive, so in particular we don't have to worry about dividing by zero. We will see later how to take care quickly of the cases where c is zero or negative.)

Note that

$$\log |g(z)| = \log |z| + (\operatorname{Re} f(z)/c) = u(z)/c,$$

and since u is between 0 and 1, we can conclude that $|g(z)|$ is between 1 and $e^{1/c}$. In other words, g maps Ω into the annulus $\mathscr{A}$ defined by

$$\mathscr{A} = \{ w \in \mathbb{C} : 1 < |w| < e^{1/c} \}.$$

To see that g covers each point of $\mathscr{A}$ precisely once, fix $w \in \mathscr{A}$. Let γ_0 and γ_1 be closed curves in Ω as in Fig. 5. More precisely, γ_0 should have winding number -1 about each point of K, and γ_0 should be close enough to the inner boundary of Ω so that each point of $g(\gamma_0)$ is smaller in absolute value than $|w|$ (see Fig. 6); this last condition can be satisfied because u is continuous on the closure of Ω and u equals 0 on the inner boundary of Ω (so $|g(z)|$ is near 1 for z near ∂K). The curve γ_1 should be chosen to have winding number one about each point of $K \cup \gamma_0$, and γ_1 should be close enough to the outer boundary of Ω so that each point of $g(\gamma_1)$ is larger in absolute value than $|w|$ (see Figs. 5 and 6); this last condition can be satisfied because u equals 1 on the outer boundary of Ω.

It is clear from Fig. 5 what we mean by the region between γ_0 and γ_1; technically, this region is defined to be the set of points about which $\gamma_0 \cup \gamma_1$ has winding number one. The argument principle states that the number of times g takes on the value w in the region between γ_0 and γ_1 equals the winding number of $g(\gamma_0) \cup g(\gamma_1)$ about w. By our choice of γ_0, the curve $g(\gamma_0)$ lies in a disk disjoint from w, and so $g(\gamma_0)$ winds zero times around w; see Fig. 6. Our construction of γ_1 shows that the disk of radius $|w|$ centered at 0 is disjoint from $g(\gamma_1)$, so the winding number of

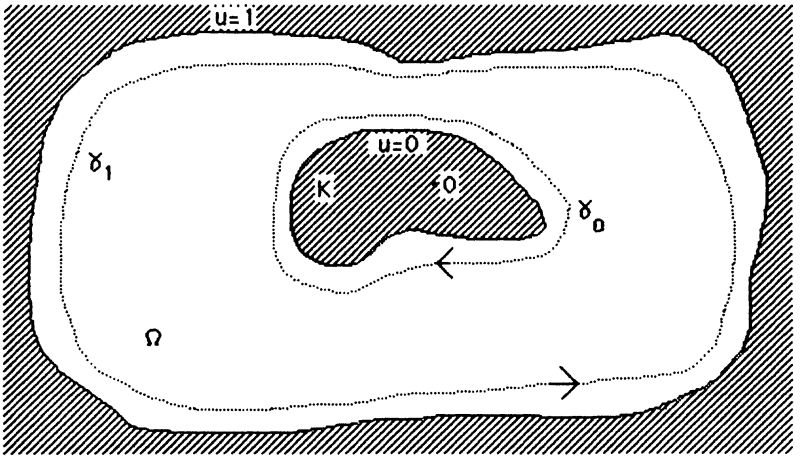


FIG. 5

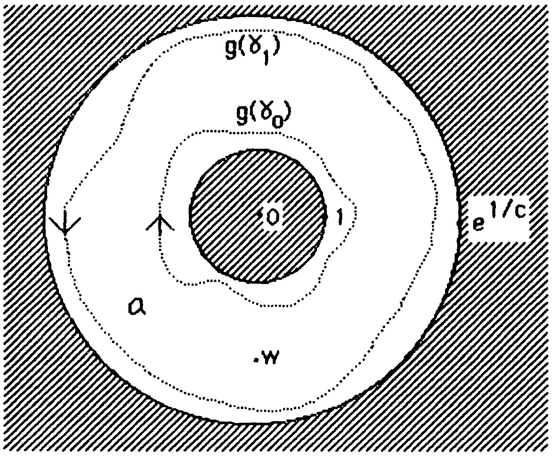


FIG. 6

$g(\gamma_1)$ about w is the same as the winding number of $g(\gamma_1)$ about 0; see Fig. 6. As usual, the winding number of $g(\gamma_1)$ about 0 is equal to the change in $(1/2\pi)\log g(z)$ as z goes once around γ_1. The definition of g shows that

$$(1/2\pi)\log g(z) = (1/2\pi)f/c + (1/2\pi)\log z.$$

Since f is analytic in Ω, the change in $(1/2\pi)f/c$ around γ_1 is zero. The change in $(1/2\pi)\log z$ around γ_1 is of course just the winding number of γ_1 around 0, and recalling that $0 \in K$ and that γ_1 winds once around each point of K, we conclude that this number equals one.

Thus g takes on the value w precisely once in the region between γ_0 and γ_1. Since we can choose γ_0 and γ_1 to be arbitrarily close to the inner and outer boundaries of Ω, we conclude that g takes on the value w precisely once on Ω. In other words, g is a one-to-one map of Ω onto the annulus $\mathscr{A}$.

Thus the proof is completed except for one minor detail that we postponed; we assumed that the constant c was positive. (The way we have defined the function u in this proof always forces c to be positive, but it is easier to deal with the other cases than to prove this.) If c is negative, simply interchange the roles of $e^{1/c}$ and 1 in defining the annulus $\mathscr{A}$ (so $e^{1/c}$ is now the inner

radius rather than the outer radius), and the above proof works fine. If c equals zero, then change the definition of g given in the proof to $g(z) = e^{f(z)}$; the same proof now shows that g maps Ω in a one-to-one fashion onto the annulus $\{w \in \mathbb{C}: 1 < |w| < e\}$. Now the proof is finished.

The applications of the Logarithmic Conjugation Theorem are far from exhausted, and I hope that those of you who attempt to find additional uses for this theorem will have as much fun with it as I have had.

References

1. Lars V. Ahlfors, Complex Analysis, 3rd ed., McGraw-Hill, New York, 1979.
2. L. Bieberbach, Conformal Mapping, Chelsea, New York, 1953.
3. Andrew Browder, Introduction to Function Algebras, Benjamin, New York, 1969.
4. John B. Conway, Functions of One Complex Variable, 2nd ed., Springer-Verlag, New York, 1978.
5. R. Courant, Dirichlet's Principle, Conformal Mapping, and Minimal Surfaces, Interscience, New York, 1950.
6. Stephen D. Fisher, Function Theory on Planar Domains, Wiley Interscience, New York, 1983.
7. G. M. Goluzin, Geometric Theory of Functions of a Complex Variable, Translations of Mathematical Monographs, vol. 26, Amer. Math. Soc., Providence, 1969.
8. Maurice Heins, Selected Topics in the Classical Theory of Functions of a Complex Variable, Holt, Rinehart and Winston, New York, 1962.
9. ______, Complex Function Theory, Academic Press, New York, 1968.
10. Yûsaku Komatu, Sur la représentation de Villat pour les fonctions analytiques définies dans un anneau circulaire concentrique, Proc. Japan Academy, 21 (1945) 94–96.
11. Henri Lebesgue, Sur les singularités des fonctions harmoniques, Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences, Paris, 176 (1923) 1097–1099.
12. Zeev Nehari, Conformal Mapping, McGraw-Hill, New York, 1952.
13. Emile Picard, Deux théorèmes élémentaires sur les singularités des fonctions harmoniques, Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences, Paris, 1976 (1923) 933–935.
14. S. Saks and A. Zygmund, Analytic Functions, 3rd ed., Elsevier Publishing, Amsterdam, and Polish Scientific Publishers, Warsaw, 1971.
15. Donald Sarason, The H^p spaces of an annulus, Memoirs of the Amer. Math. Soc., 56 (1965).
16. H. A. Schwarz, Zur Integration der partiellen Differential-gleichung $\partial^2 u/\partial x^2 + \partial^2 u/\partial y^2 = 0$, J. Reine Angew. Math., 74 (1872) 218–253.
17. H. Villat, Le problème de Dirichlet dans aire annulaire, Rend. Circ. Mat. Palermo, 33 (1912) 134–175.
18. J. L. Walsh, The approximation of harmonic functions by harmonic polynomials and by harmonic rational functions, Bull. Amer. Math. Soc., 35 (1929) 499–544.
19. Hwai-chiuan Wang, Real Hardy spaces of an annulus, Bull. Australian Math. Soc., 27 (1983) 91–105.

---

164. MISCELLANEA

Mathematical Life?

I do not remember ever having seen a sustained argument by any author which, starting from philosophical or theological premises likely to meet with general acceptance, reached the conclusion that a praiseworthy ordering of one's life is to devote it to research in mathematics.

—Sir Edmund Whittaker (1873–1956). The quotation is from *Scientific American*, Volume 183 (September 1950), page 42.

A name that is and deserves to be famous for two reasons, one of which is lifting. (See p. 275.)

He also observes that by deleting one of the sides of the Reuleaux triangle, one obtains a space Y with

$$m(Y, d^2) = \frac{1}{2}\left(3 - \sqrt{\frac{\pi}{3}}\right)$$

which is greater than the value for the Reuleaux triangle itself.

OPEN QUESTION 5. What is the value of $g_2(\mathbf{R}^n)$, the supremum of the numbers $m(X, d^2)$ as X ranges over all compact connected subsets of $\mathbf{R}^n$?

References

1. D. Amir and C. Franchetti, The radius ratio and convexity properties in normed linear spaces, Trans. Amer. Math. Soc., 282 (1984) 275–292.

2. L. M. Blumenthal and G. E. Wahlin, On the spherical surface of smallest radius enclosing a bounded subset of n-dimensional Euclidean space, Bull. Amer. Math. Soc., 47 (1941) 771–777.

3. J. Borwein and L. Keener, The Hausdorff metric and Čebyšev centres, J. Approx. Theory, 28 (1980) 366–376.

4. G. Choquet, Lectures on Analysis, vol. II, Benjamin, Reading, Massachusetts, 1969.

5. J. Cleary, Numerical Geometry, Honours project; La Trobe University, 1982 (La Trobe University Pure Mathematics Research Paper 83-2).

6. J. Cleary and S. A. Morris, Numerical geometry... not numerical topology, Math. Chronicle, 13 (1984) 47–58.

7. T. Figiel, A short proof of Dvoretzky's theorem on almost spherical sections of convex bodies, Compositio Math., 33 (1976) 297–301.

8. A. L. Garkavi, On the Chebyshev centre and convex hull of a set, Uspehi Math. Nauk, 19 (1964) no. 6. (120) 139–145.

9. O. Gross, The rendezvous value of a metric space, in Advances in game theory, Ann. of Math. Studies no. 52, Princeton, 1964, 49–53.

10. H. Jung, Ueber die kleinste Kugel, die eine räumliche Figur einschliesst, J. Reine Angew. Math., 123 (1901) 241–257.

11. S. A. Morris and P. Nickolas, On the average distance property of compact connected spaces, Arch. Math., 40 (1983) 459–463.

12. P. Nickolas, On the average distance property for subsets of the plane, to appear.

13. W. Stadje, A property of compact, connected spaces, Arch. Math., 36 (1981) 275–280.

14. J. Strantzen, An average distance result in Euclidean n-space, Bull. Austral. Math. Soc., 26 (1982) 321–330.

15. J. Stoer and C. Witzgall, Convexity and Optimization in Finite Dimensions I, Springer-Verlag, Berlin, 1970.

16. D. J. Wilson, A game with squared distance as payoff, Melbourne University Mathematics Research Report 1982, no. 16.

17. D. Yost, Average distances in compact connected spaces, Bull. Austral. Math. Soc., 26 (1982) 331–342.

---

165. MISCELLANEA

If you see someone walking about in the vicinity of a campus who looks as if he's either very backward or very brilliant, then if he's not backward he's a mathematician. There is not another group of people, I'll wager, with eccentricities so pronounced and pure, with personalities so undiluted by the attempt to conform.

—Rebecca Goldstein, *The Mind-Body Problem: a Novel*, Andre Deutsch, 1985.

ANSWER TO PHOTO ON PAGE 259

Alexandra Bellow.

NUMERICAL GEOMETRY—NUMBERS FOR SHAPES

JOAN CLEARY AND SIDNEY A. MORRIS
Mathematics Department, La Trobe University, Bundoora, Victoria, Australia, 3083

DAVID YOST
Mathematics Department, I.A.S., Australian National University, G.P.O. Box 4, Canberra, A.C.T. Australia, 2601

1. Introduction. In 1964, O. Gross [9] published a short but intriguing paper. His main result remains little known and is, in his and our experience, greeted with a range of reactions from surprise to disbelief. Before stating the result we give some concrete examples:

Let Σ^1 denote the circle (not the disk) of unit diameter and $x_1, x_2, \ldots, x_n$ any points on Σ^1. (Throughout, sets of unit diameter will be more convenient to work with than sets of unit radius.) Then there is a point y on Σ^1 such that the average distance from y to $x_1, x_2, \ldots, x_n$ is $2/\pi$. The number $2/\pi$ works for all collections of n points, for any positive integer n. Moreover, no other number will work! For the sphere Σ^2 of unit diameter in Euclidean 3-space, a similar result is true except that $2/\pi$ is replaced by $2/3$. For Σ^3 and Σ^4 the "magic numbers" are $32/15\pi$ and $72/105$, respectively. An equilateral triangle with sides of length one has "magic number" $(2+\sqrt{3})/6$, while a semicircle of unit diameter has "magic number" $4/(4+\pi)$.

Without further ado we state the Gross theorem.

THEOREM 1 [GROSS, 9]. *If (X, d) is any compact connected metric space then there is a unique positive real number $a(X, d)$ with the following property: for each positive integer n and for all (not necessarily distinct) $x_1, x_2, \ldots, x_n$ in X, there exists a y in X such that* $\frac{1}{n}\Sigma_{i=1}^{n} d(x_i, y) = a(X, d)$.

Several questions immediately present themselves:

(a) Given (X, d), how does one find $a(X, d)$?
(b) What does the number $a(X, d)$ tell us about (X, d)?
(c) What values can $a(X, d)$ take on?
(d) How does one prove the Gross theorem?
(e) How might the Gross theorem be applied to other areas of mathematics?

The first of these questions is the most tantalizing.

2. Some elementary examples. In this section we show how to calculate $a(X, d)$ in some simple cases. Our first example shows that if (X, d) is the unit interval $[0,1]$ with the usual metric,

---

Joan Cleary did her B.Sc (Hons) in Pure Mathematics at La Trobe University in 1982 with her honours project being on Numerical Geometry. She is currently a high school math teacher.

Sidney Morris is a graduate of the University of Queensland and The Flinders University of South Australia, obtaining his Ph.D. in 1970. He has held various positions at University of Queensland, University of Adelaide, University of Florida, and University of New South Wales, and Visiting Professorships at Tulane University and Tel-Aviv University. In 1974 he was a Science Research Council Visiting Fellow at the University College of North Wales. Since 1976 he has been a faculty member of La Trobe University. For $5\frac{1}{2}$ years he was Editor of the Bulletin of the Australian Mathematical Society and is currently Editor-in-Chief of the "Australian Mathematical Society Lecture Series"—a new series of books published by Cambridge University Press. He is the author of about 80 papers, most of which are on topological groups, and the book "Pontryagin Duality and the Structure of Locally Compact Abelian Groups."

David Yost is a graduate of The University of Melbourne, The Australian National University, and Edinburgh University, with his Ph.D. being conferred in 1980. He is currently a Research Fellow at The Australian National University. He has held a Queen Elizabeth II Fellowship and a La Trobe University Research Fellowship. Most of his publications are in Banach space theory.

then $a(X, d)$ exists, is unique, and equals $1/2$.

EXAMPLE 1 [**11**]. Let $X = [0,1]$ and d be the usual metric on X. We show firstly that $a(X, d)$ exists. {To do this it helps to have a peep at the answer. If $a(X, d)$ exists, then, putting $n = 2$, $x_1 = 0$, and $x_2 = 1$, we obtain for all y in $X = [0,1]$, $\frac{1}{2}\Sigma_{i=1}^{2} d(x_i, y) = \frac{1}{2}[(y - 0) + (1 - y)] = \frac{1}{2}$. So $a(X, d)$ would be unique and equal $\frac{1}{2}$.}

To see that $a(X, d)$ exists we proceed as follows: let $x_1, x_2, \dots, x_n$ be any points in X and consider the continuous "average distance" function f: $[0,1] \to [0,1]$ given by

$$f(t) = \frac{1}{n} \sum_{i=1}^{n} |x_i - t|, \quad t \in [0,1].$$

Clearly

$$f(0) = \frac{1}{n} \sum_{i=1}^{n} x_i \quad \text{and} \quad f(1) = 1 - \frac{1}{n} \sum_{i=1}^{n} x_i = 1 - f(0).$$

So either $f(0) \leqslant 1/2 \leqslant f(1)$ or $f(1) \leqslant 1/2 \leqslant f(0)$. Applying the Intermediate Value Theorem yields the existence of $y \in [0,1]$ such that $f(y) = 1/2$.

So we have shown that for any positive integer n and any $x_1, x_2, \dots, x_n$ in $[0,1]$, there exists a point y in $[0,1]$ such that $\frac{1}{n}\Sigma_{i=1}^{n} d(x_i, y) = 1/2$. Thus $1/2$ is seen to satisfy the conditions for $a(X, d)$ in Theorem 1. So existence of $a(X, d)$ is verified, and uniqueness follows from the remarks above in braces. ■

In the next three examples we assume the existence, but not the uniqueness, of $a(X, d)$. The technique used in Examples 2 and 3 should ring a bell for those familiar with game theory and, in particular, with the mini-max theorem. We shall say more about mini-max methods in §5.

EXAMPLE 2 [**5**]. Let X be the closed ball with centre O and radius $1/2$ in Euclidean m-space, $\mathbf{R}^m$, and let d be the Euclidean metric on $\mathbf{R}^m$.

Assume $a(X, d)$ exists. Firstly, let $n = 1$ and $x_1 = 0$. Then for any y in X, $d(x_1, y) \leqslant 1/2$, and so $a(X, d) \leqslant 1/2$.

Next, put $n = 2$ and choose diametrically opposite boundary points x_1' and x_2'. Then $d(x_1', x_2') = 1$, and for any y in X, $\frac{1}{2}[d(x_1', y) + d(x_2', y)] \geqslant \frac{1}{2}d(x_1', x_2') = \frac{1}{2}$. So $a(X, d) \geqslant \frac{1}{2}$. Thus $a(X, d)$ must equal $\frac{1}{2}$ (and so is unique). ■

EXAMPLE 3 [**5**]. Let X be the equilateral triangle (not its convex hull, the triangular region) with sides of length one and d the Euclidean metric. We show here that if $a(X, d)$ exists then it must equal $\dfrac{2 + \sqrt{3}}{6}$.

Firstly, let $n = 3$ and x_1, x_2, and x_3 be the vertices of the triangle. (See Fig. 1.)

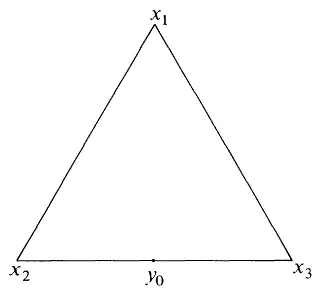


FIG. 1

It is easily verified that for $y \in X$, the quantity

$$d(x_1, y) + d(x_2, y) + d(x_3, y)$$

achieves its minimum value when y is the midpoint y_0 of any one of the sides. So for all y in X,

$$\frac{1}{3}\sum_{i=1}^{3} d(x_i, y) \geqslant \frac{1}{3}\sum_{i=1}^{3} d(x_i, y_0) = \frac{1}{3}\left(\frac{1}{2} + \frac{1}{2} + \frac{\sqrt{3}}{2}\right) = \frac{2+\sqrt{3}}{6}.$$

Thus $a(X, d) \geqslant \dfrac{2+\sqrt{3}}{6}$.

We now proceed to show that $a(X, d) \leqslant \dfrac{2+\sqrt{3}}{6}$. Again, put $n = 3$ and let x_1', x_2', and x_3' be the midpoints of the sides. (See Fig. 2.)

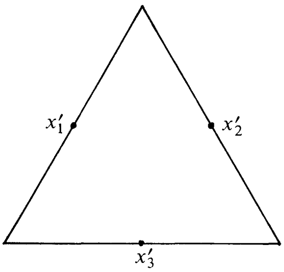


FIG. 2

The quantity $d(x_1', y) + d(x_2', y) + d(x_3', y)$ achieves its maximum value when y is any vertex y_1. So for all y in X,

$$\frac{1}{3}\sum_{i=1}^{3} d(x_i', y) \leqslant \frac{1}{3}\sum_{i=1}^{3} d(x_i', y_1) = \frac{1}{3}\left(\frac{1}{2} + \frac{1}{2} + \frac{\sqrt{3}}{2}\right) = \frac{2+\sqrt{3}}{6}.$$

Thus $a(X, d) \leqslant \dfrac{2+\sqrt{3}}{6}$.

Consequently $a(X, d) = \dfrac{2+\sqrt{3}}{6}$. ■

We point out that the same method can be used to calculate the magic number of each regular polygon in $\mathbf{R}^2$. This yields: If X_n is a regular n-gon of diameter one, then

$$a(X_n, d) = \frac{1}{2n}\sum_{k=0}^{n-1}\left[\frac{3}{2} + \frac{1}{2}\cos\frac{2\pi}{n} - \cos\frac{2k\pi}{n} - \cos\frac{2(k-1)\pi}{n}\right]^{1/2}, \quad \text{when } n \text{ is even},$$

$$a(X_n, d) = \frac{1}{n}\sum_{k=0}^{n-1}\left[\frac{\frac{3}{2} + \frac{1}{2}\cos\frac{2\pi}{n} - \cos\frac{2k\pi}{n} - \cos\frac{2(k-1)\pi}{n}}{2 - 2\cos\frac{(n-1)\pi}{n}}\right]^{1/2}, \quad \text{when } n \text{ is odd}.$$

(Full details are given in [5].) ■

OPEN QUESTION 1. If X is a general polygon in Euclidean 2-space, what is $a(X, d)$?

Incidentally, it is easy to see [5] that if X is the perimeter of a rectangle with sides of length l and w, where $l \geqslant w$, then

$$a(X, d) = \frac{1}{4}\left(l + \sqrt{4w^2 + l^2}\right)\Big/\sqrt{l^2 + w^2}.$$

EXAMPLE 4 [**11**]. Let Σ^1 be the circle of centre 0 and radius $\frac{1}{2}$ and d the Euclidean metric. Again, assume $a(\Sigma^1, d)$ exists. Let $x_0, x_1, \ldots, x_{n-1}$ be the points given by

$$x_j = \left(\frac{1}{2}\cos\frac{2\pi j}{n}, \frac{1}{2}\sin\frac{2\pi j}{n}\right), \quad j = 0, 1, \ldots, n-1,$$

which are uniformly distributed on Σ^1. Then

$$a(\Sigma^1, d) = \frac{1}{n}\sum_{j=0}^{n-1} d(x_j, y)$$

for some point y on Σ^1. By symmetry we can choose y to lie on the arc joining x_0 and x_1. So

$$d(x_0, y) \leqslant d(x_0, x_1) < \sin\frac{\pi}{n} < \frac{\pi}{n}.$$

Thus

$$\left| a(\Sigma^1, d) - \frac{1}{n}\sum_{j=0}^{n-1} d(x_j, x_0) \right| < \frac{\pi}{n}.$$

Now

$$d(x_j, x_0) = \sin\frac{\pi j}{n}.$$

Therefore

$$a(\Sigma^1, d) = \lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\sin\frac{\pi j}{n} = \int_0^1 \sin(\pi x)\,dx = \frac{2}{\pi}. \quad \blacksquare$$

3. Existence of $a(X, d)$. In Example 2, 3, and 4 we *assumed* the existence of $a(X, d)$ and proceeded to evaluate it (en route, showing its uniqueness). We give here an elementary proof of the existence of $a(X, d)$. Unfortunately, there is no elementary proof of uniqueness, and so we leave the uniqueness proof to §5.

Proof of Existence [**17**]. Let (X, d) be a compact connected metric space and put $\mathscr{F} = \bigcup_{n=1}^{\infty} X^n$. Thus $\mathscr{F}$ is the union of all ordered n-tuples with members from X. If $F \in \mathscr{F}$ is $(y_1, y_2, \ldots, y_n)$ for some integer n, we define

$$\Theta_F(x) = \frac{1}{n}\sum_{i=1}^{n} d(x, y_i).$$

[In later sections it will be more convenient to write $\Theta(x, F)$ for $\Theta_F(x)$.] As $\Theta_F\colon (X, d) \to \mathbf{R}^+$ is continuous, $\Theta_F(X)$ is a compact connected subset of $\mathbf{R}^+$. So it is a closed bounded interval; that is, $\Theta_F(X) = [a_F, b_F]$ for some $a_F, b_F \in \mathbf{R}^+$.

It is readily seen that $a(X, d)$ exists if and only if $\bigcap_{F\in\mathscr{F}}[a_F, b_F] \neq \varnothing$. We show this by verifying that for all $F, G \in \mathscr{F}$, $a_F \leqslant b_G$.

Let $G = (z_1, z_2, \ldots, z_m)$. Observe that

$$a_F \leqslant \frac{1}{n}\sum_{i=1}^{n} d(z_l, y_i), \quad \text{for each } l = 1, 2, \ldots, m,$$

and

$$\frac{1}{m}\sum_{j=1}^{m} d(y_k, z_j) \leqslant b_G, \quad \text{for each } k = 1, 2, \ldots, n;$$

it suffices to show that

$$\frac{1}{n}\sum_{i=1}^{n} d(z_l, y_i) \leqslant \frac{1}{m}\sum_{j=1}^{m} d(y_k, z_j)$$

for some $k \in \{1,2,\ldots,n\}$ and $l \in \{1,2,\ldots,m\}$.

Suppose that for all $k = 1,2,\ldots,n$ and $l = 1,2,\ldots,m$

$$\frac{1}{n}\sum_{i=1}^{n} d(z_l, y_i) > \frac{1}{m}\sum_{j=1}^{m} d(y_k, z_j),$$

and sum each side over $l = 1,2,\ldots,m$ to obtain

$$\sum_{l=1}^{m}\frac{1}{n}\sum_{i=1}^{n} d(z_l, y_i) > \sum_{j=1}^{m} d(y_k, z_j).$$

Now summing both sides over $k = 1,2,\ldots,n$ we obtain

$$\sum_{l=1}^{m}\sum_{i=1}^{n} d(z_l, y_i) > \sum_{k=1}^{n}\sum_{j=1}^{m} d(y_k, z_j).$$

But this is impossible since d is symmetric, so our supposition is false. Hence $a_F \leqslant b_G$ for all F and G in $\mathscr{F}$. Thus $a(X,d)$ exists. ∎

4. Elton's and Stadje's generalizations. In 1981 Wolfgang Stadje [**13**] published a more general result than that of Gross, although he was apparently unaware of the Gross Theorem.

THEOREM 2 [STADJE, **13**]. *If X is any compact connected Hausdorff space and f: $X \times X \to \mathbf{R}$ is a real-valued continuous symmetric function, then there is a unique real number $a(X,f)$ with the following property: for each positive integer n and for all $x_1, x_2, \ldots, x_n$ in X, there exists a y in X such that*

$$\frac{1}{n}\sum_{i=1}^{n} f(x_i, y) = a(X,f).$$

REMARK 1. The Stadje Theorem is more general than the Gross theorem since the metric d has been replaced by a continuous symmetric function f. Stadje's Theorem adds an extra "dimension" even for metric spaces, (X,d), since we can put $f = d^2$ or d^3, etc. (By d^n we mean the function given by $d^n(x,y) = (d(x,y))^n$. For sufficiently large n, d^n will not be a metric.) In fact, as we shall see later Wilson [**16**] has shown that if (X,d) is a subspace of $\mathbf{R}^2$, then it is easier to calculate $a(X,d^2)$ than $a(X,d)$. We note, incidentally, that the proof given in Gross [**9**] of his theorem easily extends to prove Stadje's Theorem. The elementary existence proof given in §3 also extends to prove the existence of $a(X,f)$. ∎

DEFINITION. If X is a compact connected Hausdorff space and f: $X \times X \to \mathbf{R}$ is a real-valued continuous symmetric function, then we put $D(X,f) = \max\{|f(x,y)|: x, y \in X\}$.

NOTATIONAL CONVENTION. From now on X will be assumed to be infinite and f will be assumed to be not the constant function zero. So $D(X,f) > 0$. (When $f = d$ is a metric, the second assumption alone implies that X is not a singleton and so, by connectedness, X must be infinite.)

We define the *magic number* or *dispersion number*, $m(X,f)$ to be $a(X,f)/D(X,f)$. The next result is a straightforward extension of a result of Gross [**9**].

PROPOSITION 1 [**5**]. *If (X,d) is any compact, connected metric space and n is any natural number, then $2^{-n} \leqslant m(X,d^n) < 1$.*

Proof. Let x_1 and x_2 be points of X such that $d(x_1,x_2) = D(X,d)$. Then Theorem 2 gives us a $y \in X$ with

$$\begin{aligned} a(X, d^n) &= \frac{1}{2}\left(d^n(x_1, y) + d^n(x_2, y)\right) \\ &\geqslant \left(\frac{1}{2}d(x_1, y) + \frac{1}{2}d(x_2, y)\right)^n \\ &\geqslant \left(\frac{1}{2}d(x_1, x_2)\right)^n = 2^{-n}D(X, d^n). \end{aligned}$$

Hence $m(X, d^{-n}) \geqslant 2^{-n}$.

That $m(X, d^n) \leqslant 1$ is clear, as $a(X, d^n)$ is an average of numbers less than or equal to $D(X, d^m)$. It remains only to show that $m(X, d^n) \neq 1$.

Suppose that $a(X, d^n) = D(X, d^n)$. Again let x_1 and x_2 be diametral points of X. Theorem 2 gives us x_3 in X such that

$$D(X, d^m) = a(X, d^n) = \frac{1}{2}\left(d^n(x_1, x_3) + d^n(x_2, x_3)\right),$$

which is the average of two numbers less than or equal to $D(X, d^n)$. Thus we must have

$$D(X, d^n) = d^n(x_1, x_3) = d^n(x_2, x_3).$$

Another application of Theorem 2 gives a point x_4 such that

$$D(X, d^n) = a(X, d^n) = \frac{1}{3}\left(d^n(x_1, x_4) + d^n(x_2, x_4) + d^n(x_3, x_4)\right).$$

As before, this forces $D(X, d^n) = d^n(x_1, x_4) = d^n(x_2, x_4) = d^n(x_3, x_4)$.

Continuing in this fashion, we inductively obtain a sequence $x_1, x_2, \dots, x_m, \dots$ such that $d^n(x_i, x_j) = D(X, d^n)$. This sequence lies in a compact metric space, but has no convergent subsequence—which is impossible. Hence our supposition was false and $a(X, d^n) \neq D(X, d^n)$. ∎

REMARK 2. It is obvious that $m(X, f)$ always lies in the closed interval $[-1, 1]$. One cannot say more than this. For example, observe that if X is any compact connected Hausdorff space, then we can define $f\colon X \times X \to \mathbf{R}$ by $f(x, y) = 1$, for all x and y in X, yielding $m(X, f) = 1$.

A more interesting example is the following: let (X, d) be any compact connected metric space, and c any point of X. Define

$$f(x, y) = d(x, y) \cdot d(c, x) \cdot d(c, y), \quad \text{for } x, y \in X.$$

Then f is not identically zero, but $m(X, f) = 0$. ∎

In his paper [**13**] Stadje also gives a formula for $a(X, f)$:

$$a(X, f) = \sup_{\mu} \inf_{\nu} \int_X \int_X f(x, y)\mu(dx)\nu(dy), \tag{1}$$

where μ and ν run through $M^1(X)$, the set of all Borel probability measures on X. We shall prove this formula in §5.

The reader unfamiliar with Borel probability measures may choose to skip from here to §6.

The formula (1) is not very useful for calculating $a(X, f)$ in general, for obvious reasons. However, Morris and Nickolas [**11**] were able to use it to calculate the magic numbers of spheres in $\mathbf{R}^n$ and topological groups. They did this by first deducing the following as a consequence of (1).

THEOREM 3 [**11**]. *Let X be a compact connected Hausdorff space and $f\colon X \times X \to \mathbf{R}$ a continuous symmetric function. If there exists a Borel probability measure μ_0 on X such that $\int_X f(x, y)\mu_0(dx)$ is independent of the choice of y in X, then $a(X, f) = \int_X f(x, y)\mu_0(dx)$ for any $y \in X$.*

As noted by Morris and Nickolas [**11**, p. 463], an application of Theorem 3 to spheres Σ^n in

Euclidean space (with μ_0 being Lebesgue measure), shows that

$$m(\Sigma^n, d) = \frac{2^{n-1}\left[\Gamma\left(\frac{n+1}{2}\right)\right]^2}{\sqrt{\pi}\,\Gamma\left(\frac{2n+1}{2}\right)},$$

where $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt$, the gamma function. (It is interesting to note that as $n \to \infty$, $m(\Sigma^n, d) \to 1/\sqrt{2}$.)

For topological groups, Morris and Nickolas [**11**] use Haar measure for μ_0.

Rather than proving Theorem 3, we note that it is a corollary of the following result of Graham Elton:

THEOREM 4. *If X is any compact connected Hausdorff space and $f\colon X \times X \to \mathbf{R}$ is a continuous symmetric function, then given any regular Borel probability measure μ on X, there is a point y in X such that $a(X, f) = \int_X f(x, y)\mu(dx)$.*

This result of Elton clearly generalizes Stadje's theorem! It will be proved in §5 once we have shown that $a(X, f)$ is unique. But first we point out another interesting application of Theorem 3.

In this example we calculate the magic number of any arc of a circle in the Euclidean plane. The method is that of Elton and depends upon Theorem 3. The "trick" is to find a measure μ_0 on the arc X such that $\int_X d(x, y)\mu_0(dx)$ is independent of the point y. [We thank Des Robbie for providing his calculations.]

EXAMPLE 5. Let X_ϕ be any arc of a circle with radius $\frac{1}{2}$ subtending an angle ϕ at the centre. Then $a(X_\phi, d) = \dfrac{4}{\phi + 4\cot\frac{\phi}{4}}$. So

$$m(X_\phi, d) = \begin{cases} \dfrac{4}{\phi + 4\cot\frac{\phi}{4}}, & \text{if } \pi \leqslant \phi \leqslant 2\pi, \\[2ex] \dfrac{4}{\phi\sin\frac{\phi}{2} + 8\cos^2\frac{\phi}{4}}, & \text{if } 0 < \phi < \pi. \end{cases}$$

Proof. Let $P_\theta \in X_\phi$ be the point $\left(\frac{1}{2}\cos\theta, \frac{1}{2}\sin\theta\right)$, $0 \leqslant \theta \leqslant \phi$. Put $\mu_0 = \mu_1 + \mu_2 + \mu_3$, where μ_1 is 1-dimensional measure on the arc of the circle, μ_2 is a point measure giving weight m to P_0, and μ_3 is a point measure giving weight m to P_ϕ. Then $\mu_0\Big/\left(\frac{\phi}{2} + 2m\right)$ is a probability measure on X_ϕ. (See Fig. 3.)

By Theorem 4 there is a point $P_\alpha \in X_\phi$ such that

$$a(X_\phi, d) = \frac{\int_0^\phi d(P_\alpha, P_\theta)\left(d\frac{1}{2}\theta\right) + m\left(\sin\frac{\alpha}{2} + \sin\frac{\phi - \alpha}{2}\right)}{\frac{\phi}{2} + 2m}.$$

So

$$a(X_\phi, d) = \frac{2 + \left[m\sin\frac{\phi}{2} - \cos\frac{\phi}{2} - 1\right]\cos\frac{\alpha}{2} + \left[m - m\cos\frac{\phi}{2} - \sin\frac{\phi}{2}\right]\sin\frac{\alpha}{2}}{\frac{\phi}{2} + 2m}.$$

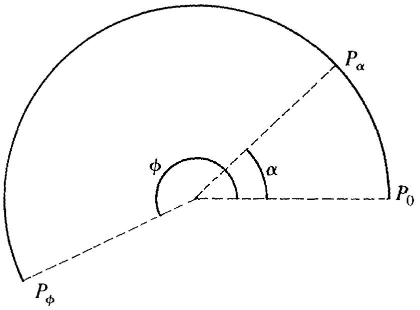


FIG. 3

Now, the term

$$\left(m\sin\frac{\phi}{2}-\cos\frac{\phi}{2}-1\right)\cos\frac{\alpha}{2}+\left(m-m\cos\frac{\phi}{2}-\sin\frac{\phi}{2}\right)\sin\frac{\alpha}{2}$$

will be independent of α if

$$\left(m\sin\frac{\phi}{2}-\cos\frac{\phi}{2}-1\right)=\left(m-m\cos\frac{\phi}{2}-\sin\frac{\phi}{2}\right)=0;$$

that is, when

$$m=\frac{\sin\frac{\phi}{2}}{1-\cos\frac{\phi}{2}}$$

So we choose μ_2, μ_3 to be the point measures which give weight

$$m=\frac{\sin\frac{\phi}{2}}{1-\cos\frac{\phi}{2}}$$

to each endpoint. Then

$$a(X_\phi, d)=\frac{4}{\phi+4\cot\frac{\phi}{4}}.$$

When $0<\phi<\pi$, the diameter of X_ϕ, $D(X_\phi, d)=\sin\frac{\phi}{2}$. So

$$m(X_\phi, d)=\frac{4}{\phi\sin\frac{\phi}{2}+8\cos^2\frac{\phi}{4}},\quad \text{for } 0<\phi<\pi.$$

When $\pi\leqslant\phi\leqslant 2\pi$, $D(X_\phi, d)=1$, and therefore

$$m(X_\phi, d)=\frac{4}{\phi+4\cot\frac{\phi}{4}}.$$

(So if $\phi=2\pi$, $m(X_\phi, d)=m(\Sigma^1, d)=2/\pi$, which agrees with Example 4.) ∎

While we have calculated the magic number of the circle, sphere, and arc of a circle, we do not know how to find the magic number of an ellipse.

OPEN QUESTION 2. What is the magic number of a general ellipse?

5. Uniqueness of $a(X,f)$. In §3 we proved the existence of $a(X,d)$ for any compact connected metric space (X,d) and in §4 indicated that this proof carried over to $a(X,f)$. It is now time to prove the uniqueness of $a(X,f)$. But first, it will be convenient to prove the existence part of Elton's Theorem 4.

THEOREM 4′. *Let X and f be as in the statement of Theorem 4. Then there exists a real number a such that, given any regular Borel probability measure μ on X, there is a point $x \in X$ with $a = \int_X f(x,y)\mu(dy)$.*

Proof. Equipped with the weak*-topology, $M^1(X)$, the set of all regular Borel probability measures on X, becomes a compact convex set. Its extreme points are the point masses corresponding to elements of X. So elements of $\mathscr{F}$, the set of all ordered n-tuples from X, may be regarded as convex combinations, with rational coefficients, of extreme points of $M^1(X)$. From the Krein-Milman theorem, we deduce that $\mathscr{F}$ is dense in $M^1(X)$. (See Choquet [4, §25].)

Define $\Theta\colon X \times M^1(X) \to \mathbf{R}$ by $\Theta(x,\mu) = \int_X f(x,y)\mu(dy)$. (The number $\Theta(x,\mu)$ represents the distance from the point x to the probability measure μ.) It is routine to show that Θ is continuous. For $\mu \in M^1(X)$ let

$$a(\mu) = \min\{\Theta(x,\mu)\colon x \in X\} \quad \text{and} \quad b(\mu) = \max\{\Theta(x,\mu)\colon x \in X\}.$$

As in the previous existence proof, we have $\{\Theta(x,\mu)\colon x \in X\} = [a(\mu), b(\mu)]$. The conclusion of this theorem may be stated as

$$(\exists a \in \mathbf{R})(\forall \mu \in M^1(X))\, a \in [a(\mu), b(\mu)].$$

As before, this can be the case only if $a(\mu) \leqslant b(\nu)$ for all $\mu, \nu \in M^1(X)$. Since $a_F \leqslant b_G$ for all F, $G \in \mathscr{F}$, the result follows from the density of $\mathscr{F}$ in $M^1(X)$. ∎

Proof of uniqueness of $a(X,f)$. We continue the notation from the previous proof. That proof shows that uniqueness of $a(X,f)$ is just the statement $\sup_{F\in\mathscr{F}} a_F = \inf_{G\in\mathscr{F}} b_G$, and that $a(X,f)$ is their common value. From continuity of Θ, it follows that $a, b\colon M^1(X) \to \mathbf{R}$ are continuous functions. From the density of $\mathscr{F}$, it follows that

$$\sup_{F\in\mathscr{F}} a_F = \max_{\mu\in M^1(X)} a(\mu) \quad \text{and} \quad \inf_{G\in\mathscr{F}} b_G = \min_{\nu\in M^1(X)} b(\nu).$$

Now for $\mu, \nu \in M^1(X)$, let

$$A(\mu,\nu) = \int_X\int_X f(x,y)\mu(dy)\nu(dx).$$

For fixed μ, A is an affine continuous function of ν. By Bauer's maximum principle, [4, §25] it attains its minimum at an extreme point of $M^1(X)$. Thus

$$\min_\nu A(\mu,\nu) = \min_x \Theta(x,\mu) = a(\mu).$$

Similarly

$$\max_\mu A(\mu,\nu) = \max_y \Theta(y,\nu) = b(\nu),$$

and the uniqueness statement becomes

$$\max_\mu \min_\nu A(\mu,\nu) = \min_\nu \max_\mu A(\mu,\nu). \tag{2}$$

This is just Fan's version of the mini-max theorem [15, 6.3.8]. So uniqueness is proved. Also observe, now, that formula (2) yields formula (1). ∎

REMARK 3. It is now clear that uniqueness of $a(X,f)$ is essentially equivalent to the mini-max theorem, whereas existence of $a(X,f)$ corresponds to the trivial mini-max inequality. This is why existence of $a(X,f)$ is easier to prove than uniqueness. ∎

Reexamining Examples 2 and 3 we observe that the technique used was to find measures μ and

ν which maximized $a(\mu)$ and minimized $b(\nu)$. Since the resultant $a(\mu)$ and $b(\nu)$ are equal, their common value is $a(X, d)$. This is the same as finding μ and ν for which

$$\Theta(x,\nu) \leqslant a(X,f) \leqslant \Theta(x,\mu), \quad \text{for all } x \in X.$$

The above uniqueness proof shows that it is always possible to do this. (These μ and ν are called "optimal strategies".)

A question that arises naturally is: can we choose $\mu = \nu$? If so, we would have a measure μ for which $\Theta(x, \mu) = a(X, f)$ for all $x \in X$. This was essentially the idea used to calculate $a(X, d)$ in Examples 4 and 5. The next result shows that it is not always possible to find such a probability measure. In particular, it is not possible in Examples 2 and 3. The idea of the following proof is due to David Wilson.

PROPOSITION 2. *Let X be a compact connected subset of a rotund normed space. (This means that $\|x + y\| < \|x\| + \|y\|$ unless x and y are linearly dependent. Observe that $\mathbf{R}^n$, for any n, is a rotund normed space.) Suppose that, for some probability measure μ on X, $\int_X \|x - y\|\mu(dy) = \Theta(x, \mu)$ is independent of x, and that X is not a line segment. Then no three points of X are collinear.*

Proof. Suppose that, for some line L, $X \cap L$ contains at least three points. We show that X must be a line segment. Let a, b be two points in $X \cap L$ which are as far apart as possible, and let c be any other point in $X \cap L$. Then $c = \lambda a + (1 - \lambda)b$ for some $\lambda \in (0,1)$. By rotundity,

$$\|x - c\| < \lambda\|x - a\| + (1 - \lambda)\|x - b\|,$$

for all $x \in X \setminus \{a, b\}$. Suppose μ is not concentrated on $\{a, b\}$. Integrating over X, we obtain

$$\Theta(c,\mu) < \lambda\Theta(a,\mu) + (1-\lambda)\Theta(b,\mu).$$

This contradicts our assumption that $\Theta(x, \mu)$ is independent of x. Thus μ is concentrated on $\{a, b\}$.

So

$$\Theta(x,\mu) = \mu\{a\}\|x - a\| + \mu\{b\}\|x - b\|$$

for all $x \in X$. Since $\Theta(a, \mu) = \Theta(b, \mu)$, we must have $\mu\{a\} = \mu\{b\} = \frac{1}{2}$. Then, for any $x \in X$, we have

$$\|x - a\| + \|x - b\| = 2\Theta(x,\mu) = 2\Theta(a,\mu) = \|a - b\|.$$

By rotundity, $x - a$ and $x - b$ are linearly dependent. Thus X lies in the line segment

$$[a,b] = \{\alpha a + (1-\alpha)b : 0 \leqslant \alpha \leqslant 1\}.$$

Finally, connectedness forces $X = [a, b]$. ∎

6. The range of $m(X,d)$. Often it is difficult to calculate $a(X, f)$ precisely, and we must be content with some sort of estimate. In such cases, the following observation is surprisingly useful.

LEMMA 1 [17]. *Suppose, for fixed α, $\beta \in \mathbf{R}$, that (X, f) has the following property: given $F \in \mathscr{F}$, there is a point $x \in X$ with $\alpha \leqslant \Theta(x, F) \leqslant \beta$. Then $\alpha \leqslant a(X, f) \leqslant \beta$.*

Lemma 1 has already been used in some of the previous examples. Two further examples are given below.

EXAMPLE 6. Let (X, d) be the subspace of $\mathbf{R}^2$ given by

$$X = \{(y,z) : y, z \in \mathbf{R}, \gamma^2 \leqslant y^2 + z^2 \leqslant \delta^2\},$$

where γ and δ are positive real numbers. (See Fig. 4.)

Let $p = (0, \gamma)$. Then for all $x \in X$, $d(x, p) \leqslant \gamma + \delta$. Hence, by Lemma 1 $a(X, d) \leqslant \gamma + \delta$. ∎

EXAMPLE 7. Let $X = \left\{\left(x, \sin\dfrac{1}{x}\right) : 0 < x \leqslant \dfrac{1}{\pi}\right\} \cup \{(0, y) : y \in [-1, 1]\}$. Then X is a com-

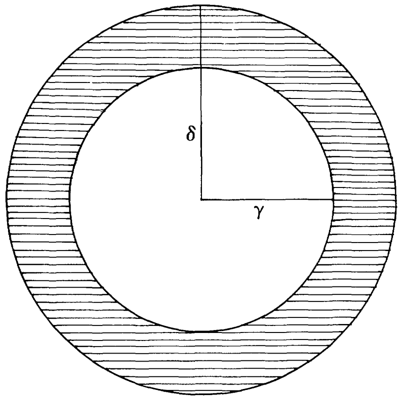


FIG. 4

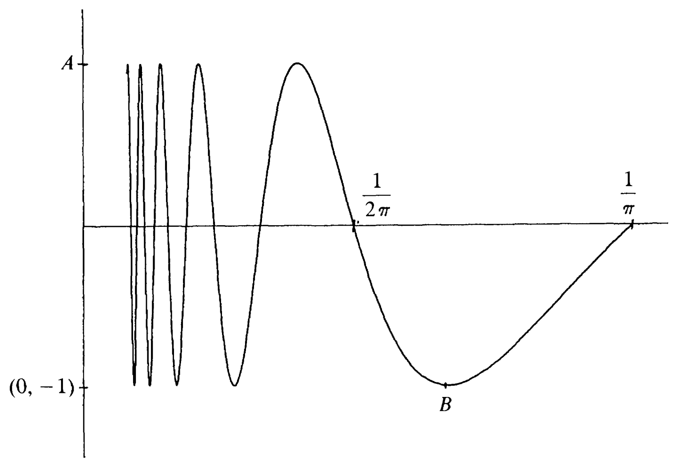


FIG. 5

pact connected subset of $\mathbf{R}^2$.

Let $x_1 = \left(\frac{1}{2\pi}, 0\right)$, $A = (0,1)$, $B = \left(\frac{2}{3\pi}, -1\right)$. (See Fig. 5.)

For all $y \in X$,

$$\begin{aligned} d(x_1, y) &\leqslant d(x_1, A) \\ &= \left[1 + \frac{1}{4\pi^2}\right]^{1/2}. \end{aligned}$$

By Lemma 1

$$a(X,d) \leqslant \frac{1}{2\pi}[4\pi^2 + 1]^{1/2}.$$

Also

$$\begin{aligned} D(X,d) \geqslant d(A,B) &= \left[4 + \frac{4}{9\pi^2}\right]^{1/2} \\ &= \frac{2}{3\pi}[9\pi^2 + 1]^{1/2}. \end{aligned}$$

Hence

$$m(X,d) = \frac{a(X,d)}{D(X,d)} \leqslant \frac{\frac{1}{2\pi}[4\pi^2+1]^{1/2}}{\frac{2}{3\pi}[9\pi^2+1]^{1/2}}$$

$$\simeq 0.5035.$$

So $0.500 \leqslant m(X,d) \leqslant 0.504$. ■

We shall now restrict our attention to the metric case, and investigate the range of values possible for $m(X,d)$, when various restrictions are placed on (X,d).

Gross [**9**] showed that $m(X,d)$ always lies in the half-open interval $\left[\frac{1}{2},1\right)$. This is a special case of Proposition 1. If $m(X,d)$ is close to 1, this indicates that "most" points of X are far apart from one another. On the other hand, it is possible to have $m(X,d)=\frac{1}{2}$ for a space in which "many" points are far apart from one another. As an example of this phenomenon, observe that the magic number for any circle is $2/\pi$ (Example 4), whereas the magic number for a circle together with any of its diameters is $\frac{1}{2}$ (use the method of Example 2). More generally, the following is true.

PROPOSITION 3. *Let (X,d) be any compact connected metric space, and choose m with $\frac{1}{2} \leqslant m < m(X,d)$. Then there exists a wedge $X \vee I$ (that is, a space obtained from X by gluing on a line segment I) such that the magic number of $X \vee I$ equals m.*

Proposition 3 was first proved in [**17**], but a simpler proof can be found in [**6**].

NOTATION. For each metric space (S,d), we define $g(S)$ to be the supremum of the numbers $m(X,d)$ as X ranges over all compact connected subsets of S.

If the supremum is attained, then, by Proposition 1, it is strictly less than one. This is the case, for example, for any finite-dimensional normed space [**17**, Theorem 5]. (If E is an infinite dimensional normed space, then $g(E)=1$, by Dvoretzky's theorem [**7**] and [**17**, Theorem 9]. In this case, by Proposition 1, the supremum cannot be attained.)

OPEN QUESTION 3. Let $\mathbf{R}^n$ denote Euclidean n-space. Then what is $g(\mathbf{R}^n)$? Indeed what is $g(\mathbf{R}^2)$?

Gross [**9**] stated that "computations seem to indicate that the bound (that is, $g(\mathbf{R}^2)$) is not much greater than 2/3". We suggest that $g(\mathbf{R}^2)$ may equal the magic number of the Reuleaux triangle. (The Reuleaux triangle consists of the vertices of an equilateral triangle together with three arcs of circles, each circle having centre at one of the vertices and endpoints, the other two vertices. See Fig. 6.)

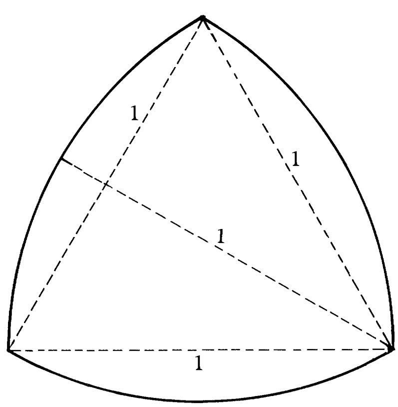


FIG. 6

The magic number for the Reuleaux triangle is not known exactly, but computer approximation shows that it is about 0.668.

OPEN QUESTION 4. What is the magic number of the Reuleaux triangle?

In 1982 Graham Elton (private communication, see also [5]) proved that $g(\mathbf{R}^2) \leqslant 0.775$. More recently Peter Nickolas [12] has proved that $g(\mathbf{R}^2) \leqslant 4/\pi\sqrt{3} \simeq 0.735$. From this it follows that $0.667 < g(\mathbf{R}^2) < 0.736$. Nickolas does this by proving that if (X, d) is any compact connected metric subspace of a disk of diameter one in $\mathbf{R}^2$, then $a(X, d) \leqslant 2/\pi$, which is, of course, the magic number of the circle of diameter one. Then, using the fact that every subset of $\mathbf{R}^2$ of diameter not greater than one is a subset of a disk of diameter $2/\sqrt{3}$, one obtains that for any compact connected metric subspace (Y, d) of $\mathbf{R}^2$

$$m(Y, d) \leqslant \frac{2}{\pi} \cdot \frac{2}{\sqrt{3}} = \frac{4}{\pi\sqrt{3}}.$$

Using a similar technique, but with the disk replaced by a hexagon, the third author has recently improved this estimate to $g(\mathbf{R}^2) \leqslant \dfrac{2+\sqrt{3}}{3\sqrt{3}} \simeq 0.718$. This uses the fact that any set in $\mathbf{R}^2$ of diameter one is contained in a regular hexagon of side length $1/\sqrt{3}$.

While we do not know the exact value of $g(\mathbf{R}^2)$, the analogous problem for *convex* subsets of $\mathbf{R}^2$ (indeed $\mathbf{R}^n$) has been solved. To attack this problem, it is appropriate to consider convex sets in general normed linear spaces.

NOTATION. Let X be a compact convex subset of some normed space. We denote by $r(X)$ the minimum of the radii of those closed balls, with centres in X, which contain X. A simple compactness argument shows that this minimum exists.

It is essential in this definition to consider only balls with centres in X. A result of Garkavi [8] asserts that if E is a normed space of dimension at least three, which does not admit an inner-product structure, then E contains a compact convex set X, which is contained in a ball of radius strictly less then $r(X)$. See Borwein and Keener [3].

The following elegant result is due to Esther and George Szekeres (private communication).

THEOREM 5. *Let X be a compact convex subset of some normed space, with d being the metric given by the norm. Then $a(X, d) = r(X)$.*

Proof (Szekeres). Let x_0 be the centre of a ball, of radius $r = r(X)$, which contains X. Then $d(x, x_0) \leqslant r$ for all $x \in X$. Choose $x_1, \dots, x_n \in X$ and define Θ: $X \to \mathbf{R}$ by

$$\Theta(x) = \frac{1}{n} \sum_{i=1}^{n} d(x, x_i).$$

We must show that Θ takes the value r somewhere on X. Clearly $\Theta(x_0) \leqslant r$. Since X is connected, it suffices to show that $\Theta(x) \geqslant r$ for some $x \in X$.

We let $c = \frac{1}{n}\sum_{i=1}^{n} x_i \in X$, and choose $y \in X$ so that $d(c, y) \geqslant r$. At least one such y exists, by definition of $r(X)$. Then

$$r \leqslant \|c - y\| \leqslant \frac{1}{n} \sum_{i=1}^{n} \|x_i - y\| = \Theta(y),$$

as required. ■

We remark that this proof does not use the mini-max technique so common in previous examples.

We can now calculate the convex analogue of $g(\mathbf{R}^n)$.

NOTATION. For any normed space E, let $s(E)$ be the supremum of the numbers $m(X, d)$ as X

ranges over all compact convex subsets of E.

THEOREM 6 [14]. *For each positive integer n, $s(\mathbf{R}^n) = \sqrt{n/(2n+2)}$.*

Proof. A classical result of Jung [10] states that if X is a compact convex subset of $\mathbf{R}^n$, with unit diameter, then $r(X) \leqslant \sqrt{n/(2n+2)}$. Blumenthal and Wahlin [2] gave a simple proof of this, and also showed that the upper bound is attained when X is the regular n-simplex. The result now follows from Theorem 5. ∎

For recent work relating to Jung's theorem, we refer the reader to [1], and the references therein.

REMARK 4. The proof of Theorem 6 first given by Strantzen [14] is direct, and independent of Theorem 5. He also calculated the magic numbers for the k-skeletons of a regular n-simplex, $1 \leqslant k \leqslant n$. (The k-skeleton of an n-simplex X is the union of all k-simplices whose vertices are already vertices of X.) Stadje [13] proved that $s(\mathbf{R}^2) \leqslant \frac{1}{2}\sqrt{5 - 2\sqrt{3}}$, a weaker result than Theorem 6. He claimed that the same bound was valid for $s(\mathbf{R}^n)$, but, as shown by Strantzen [14], this is false for $n \geqslant 4$.

REMARK 5. Note that as n tends to infinity, $s(\mathbf{R}^n)$ tends to $1/\sqrt{2}$.

REMARK 6. It is shown in [17] that the maximum of $s(E)$ as E ranges over all n-dimensional normed spaces, is $n/(n+1)$.

7. Numerical geometry... not numerical topology. In §6 we remarked that the magic number provides some information on how "spread out" a space is. The following surprising result shows that the magic number depends heavily on the metric, not simply on the topology of the space. So it is appropriate to call this subject numerical geometry, rather than numerical topology.

THEOREM 7 [6]. *Let X be a compact connected metrizable space. Then for each real $m \in [\frac{1}{2}, 1)$, there is a compatible metric d on X such that $m(X, d) = m$.*

Outline Proof. Choose a compatible metric ρ on X so that $D(X, \rho) = 1$ and then choose $a, b, c \in X$ with $\rho(a, b) = 1$ and $\rho(a, c) = \frac{1}{2}$. Then define d by

$$d(x, y) = \min\left\{\rho(x, y), \min\left\{\frac{1}{2}, \rho(c, x)\right\} + \min\left\{\frac{1}{2}, \rho(c, y)\right\}\right\}.$$

It can be verified that d is a metric, equivalent to ρ, and that $m(X, d) = \frac{1}{2}$.

For the case $m > \frac{1}{2}$, consider the family of metrics $d_\lambda (\lambda \geqslant 0)$ defined by

$$d_\lambda = (\lambda + 1) d/(\lambda d + 1).$$

Then $D(X, d_\lambda) = 1$ for all λ, and if $x \neq y$, then $d_\lambda(x, y) \to 1$ as $\lambda \to \infty$. Some straightforward but tedious analysis then shows that $m(X, d_\lambda)$ is a continuous function of λ, and that $m(X, d_\lambda) \to 1$ as $\lambda \to \infty$. Since $m(X, d_0) = \frac{1}{2}$, it follows from the Intermediate Value Theorem that $m(X, d_\lambda) = m$ for some λ. ∎

For compact connected Hausdorff spaces, the following analogue of Theorem 7 holds. We give a new proof.

THEOREM 8 [6]. *Let X be a compact connected Hausdorff space. Then for each $m \in [-1, 1]$ there is a continuous symmetric function f: $X \times X \to \mathbf{R}$ such that $m(X, f) = m$.*

Proof. Firstly let $m \in [0, 1]$. Choose distinct points a, b, c in X and put

$$S = (X \times \{c\}) \cup (\{c\} \times X) \cup \{(a, b)\} \cup \{(b, a)\}.$$

So S is a closed subspace of the compact Hausdorff space $X \times X$. Define a continuous function g: $S \to [0, 1]$ as follows: $g(x, c) = g(c, x) = \sqrt{m}$, for all $x \in X$, and $g(a, b) = g(b, a) = 1$. By

Tietze's extension theorem, there exists a continuous function θ: $X \times X \to [0,1]$ such that $g(s) = \theta(s)$, for all $s \in S$. Now define f: $X \times X \to [0,1]$ by $f(x, y) = \theta(x, y)\theta(y, x)$, for all $(x, y) \in X \times X$. Then f is a continuous symmetric function with the property that $f(x, c) = m$, for all $x \in X$. From this it immediately follows that $a(X, f) = m$. As $D(X, f) = 1$, we have $m(X, f) = m$, as required.

If $m \in [-1, 0]$, then we find, as above, a function f such that $m(X, f) = -m$. Then putting $f' = -f$, we have $m(X, f') = m$. ■

8. The squared distance function. As indicated in Remark 1 the calculation of $a(X, d^2)$ for subspaces (X, d) of Euclidean space, $\mathbf{R}^n$, is much simpler than that of $a(X, d)$. We shall only touch on the topic here as full details appear in the paper **[16]** of David Wilson. The key theorem is:

THEOREM 9. *Let X be a compact connected subset of $\mathbf{R}^n$. Let B_1 be a closed ball and B_2 an open ball such that $X \subset B_1 \setminus B_2$ and the centre of each ball lies in the closed convex hull of the intersection of X with the boundary of the other. Further, let B_1 have centre α and radius R and let B_2 have centre β and radius r. Then $a(X, d^2) = R^2 + r^2 - |\alpha - \beta|^2$.*

If we consider compact *convex* subsets X of $\mathbf{R}^n$, then by the hypotheses of Theorem 9, r must be zero and α must equal β. Therefore $a(X, d^2) = R^2$. Using a modification of the Szekeres proof (Theorem 5) that the magic number of a compact convex set is the circumradius of the set, Wilson shows that R is the circumradius of X.

Open Question 2 asks for the magic number of an ellipse. The analogous problem for the squared distance function is easy using Theorem 9.

EXAMPLE 8 **[16]**. Let X be the ellipse with semi-major axis of length R and semi-minor axis of length r. Clearly $\alpha = \beta$, and so $a(X, d^2) = R^2 + r^2$. (See Fig. 7.) ■

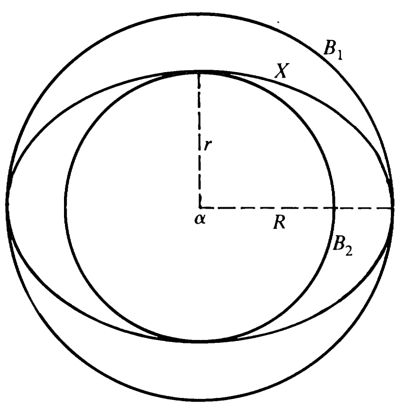


FIG. 7

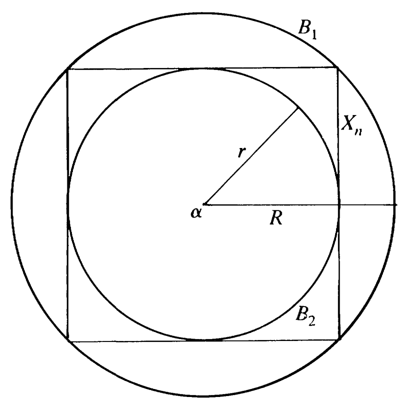


FIG. 8

The regular n-gon is also easily handled:

EXAMPLE 9 **[16]**. Let X_n be the regular n-sided polygon with vertices on the circle of radius $\frac{1}{2}$ and centre the origin. (See Fig. 8.)

Again $\alpha = \beta$. Now $R^2 = \frac{1}{4}$ and $r^2 = \frac{1}{8}\left(1 + \cos\dfrac{2\pi}{n}\right)$. Therefore $a(X, d^2) = \frac{1}{8}\left(3 + \cos\dfrac{2\pi}{n}\right)$.
■

Open Question 4 asks for the magic number of the Reuleaux triangle. Wilson **[16]** shows that

$$m(X, d^2) = \frac{5 - 2\sqrt{3}}{3}.$$

He also observes that by deleting one of the sides of the Reuleaux triangle, one obtains a space Y with

$$m(Y, d^2) = \frac{1}{2}\left(3 - \sqrt{\frac{\pi}{3}}\right)$$

which is greater than the value for the Reuleaux triangle itself.

OPEN QUESTION 5. What is the value of $g_2(\mathbf{R}^n)$, the supremum of the numbers $m(X, d^2)$ as X ranges over all compact connected subsets of $\mathbf{R}^n$?

References

1. D. Amir and C. Franchetti, The radius ratio and convexity properties in normed linear spaces, Trans. Amer. Math. Soc., 282 (1984) 275–292.

2. L. M. Blumenthal and G. E. Wahlin, On the spherical surface of smallest radius enclosing a bounded subset of n-dimensional Euclidean space, Bull. Amer. Math. Soc., 47 (1941) 771–777.

3. J. Borwein and L. Keener, The Hausdorff metric and Čebyšev centres, J. Approx. Theory, 28 (1980) 366–376.

4. G. Choquet, Lectures on Analysis, vol. II, Benjamin, Reading, Massachusetts, 1969.

5. J. Cleary, Numerical Geometry, Honours project; La Trobe University, 1982 (La Trobe University Pure Mathematics Research Paper 83-2).

6. J. Cleary and S. A. Morris, Numerical geometry... not numerical topology, Math. Chronicle, 13 (1984) 47–58.

7. T. Figiel, A short proof of Dvoretzky's theorem on almost spherical sections of convex bodies, Compositio Math., 33 (1976) 297–301.

8. A. L. Garkavi, On the Chebyshev centre and convex hull of a set, Uspehi Math. Nauk, 19 (1964) no. 6. (120) 139–145.

9. O. Gross, The rendezvous value of a metric space, in Advances in game theory, Ann. of Math. Studies no. 52, Princeton, 1964, 49–53.

10. H. Jung, Ueber die kleinste Kugel, die eine räumliche Figur einschliesst, J. Reine Angew. Math., 123 (1901) 241–257.

11. S. A. Morris and P. Nickolas, On the average distance property of compact connected spaces, Arch. Math., 40 (1983) 459–463.

12. P. Nickolas, On the average distance property for subsets of the plane, to appear.

13. W. Stadje, A property of compact, connected spaces, Arch. Math., 36 (1981) 275–280.

14. J. Strantzen, An average distance result in Euclidean n-space, Bull. Austral. Math. Soc., 26 (1982) 321–330.

15. J. Stoer and C. Witzgall, Convexity and Optimization in Finite Dimensions I, Springer-Verlag, Berlin, 1970.

16. D. J. Wilson, A game with squared distance as payoff, Melbourne University Mathematics Research Report 1982, no. 16.

17. D. Yost, Average distances in compact connected spaces, Bull. Austral. Math. Soc., 26 (1982) 331–342.

---

165. MISCELLANEA

If you see someone walking about in the vicinity of a campus who looks as if he's either very backward or very brilliant, then if he's not backward he's a mathematician. There is not another group of people, I'll wager, with eccentricities so pronounced and pure, with personalities so undiluted by the attempt to conform.

—Rebecca Goldstein, *The Mind-Body Problem: a Novel*, Andre Deutsch, 1985.

ANSWER TO PHOTO ON PAGE 259

Alexandra Bellow.

SUMS OF THREE SQUARES AND LEVELS OF QUADRATIC NUMBER FIELDS

CHARLES SMALL*

Department of Mathematics and Statistics, Queen's University, Kingston, Ontario, Canada K7L 3N6

A famous theorem, attributed in [S] to Gauss, characterizes integers which can be written as sums of three squares:

THEOREM 1. *Let n be a positive integer and write $n = 4^a n', 4 \nmid n', a \geqslant 0$. Then* $n = \boxed{3} \Leftrightarrow n' \not\equiv 7 \pmod 8$.

Here and in what follows we write $\boxed{m}$ for a sum of m squares. $\mathbb{Z}$ and $\mathbb{Q}$ will denote respectively the integers and the rational numbers.

In [S] Gauss' theorem is shown to follow from the standard facts about quadratic forms over the p-adic fields $\mathbb{Q}_p$. In a recent paper [G] the same theorem is derived from the standard facts about splitting fields for the quaternions over $\mathbb{Q}$. And in [R] a beautiful connection is observed between Gauss' theorem and the level of quadratic number fields.

Our aim in this note is to elaborate the result of [R], giving simple direct proofs, and thereby make Gauss' theorem available as an appealing application in courses where quadratic number fields are discussed, in much the same way as [G] makes it available for courses which deal with central simple algebras. Our approach is similar in spirit to [G], although the machinery we appeal to is more elementary.

To begin, we note two consequences of Theorem 1:

COROLLARY 2. *Let n be a squarefree positive integer; then* $n = \boxed{3} \Leftrightarrow n \not\equiv 7 \pmod 8$.

Proof. This is just a special case of Theorem 1, since $4^a n'$ is squarefree only when $a = 0$.

COROLLARY 3 (Lagrange, 1770). *Let n be any positive integer; then* $n = \boxed{4}$.

Proof. Write $n = 4^a n', 4 \nmid n', a \geqslant 0$. If $n' \not\equiv 7 \pmod 8$, then $n = \boxed{3}$, so $n = \boxed{4}$. If $n' \equiv 7 \pmod 8$, then $n' - 1 = \boxed{3}$, so $n' = \boxed{3} + 1$ and therefore $n = \boxed{4}$. (Notice this shows slightly more than $n = \boxed{4}$: n is either $\boxed{3}$ or $\boxed{3} + 4^a$ for some $a \geqslant 0$.)

An important step in virtually all approaches to Gauss' theorem is the following fact, referred to in [S] as a special case of the Davenport-Cassels lemma:

LEMMA 4. *Let $n \in \mathbb{Z}$; then* $n = \boxed{3}$ *in* $\mathbb{Q} \Leftrightarrow n = \boxed{3}$ *in* $\mathbb{Z}$.

Of course, the implication " $\Leftarrow$ " is trivial. The converse, which says that if an integer is the sum of three rational squares then it is in fact the sum of three integer squares, is efficiently proved in [S], and essentially the same proof is offered in [G], so we will not repeat it here.

Charles Small received his doctorate in 1969, under Professor Hyman Bass, at Columbia University. He has been at Queen's University since 1970 except for leaves in 1977–78 and 1983–85, and has published more than a baker's dozen papers in algebra and number theory. He is also co-author of two books, one on the Brauer group of a commutative ring and one on homological methods in commutative algebra.

*Research supported, in part, by NSERC Canada.

Using Lemma 4, we show next that Theorem 1 is actually equivalent to its special case Corollary 2. That is, assuming Corollary 2, we prove Theorem 1. Indeed, given a positive integer n, first split off the squarefree part: $n = 4^a n'$, $4 \nmid n'$, $a \geqslant 0$, $n' = km$, k an odd square (possibly 1), m squarefree. Observe that $n' \equiv m \pmod 8$, since any odd square is $\equiv 1 \pmod 8$. Then by Lemma 4 we have that $n = \boxed{3}$ in $\mathbb{Z} \Leftrightarrow m = \boxed{3}$ in $\mathbb{Z}$, and by Corollary 2, $m = \boxed{3}$ in $\mathbb{Z} \Leftrightarrow m \not\equiv 7 \pmod 8 \Leftrightarrow n' \not\equiv 7 \pmod 8$. Thus, given Lemma 4, Theorem 1 and its Corollary 2 are equivalent.

Next we introduce the *level* ("Stufe" in German, where it originally appeared) of a field and state the theorem which computes it for quadratic number fields. The level $s(\mathbf{K})$ of a field $\mathbf{K}$ is by definition the smallest $n \geqslant 1$ such that $-1 = \boxed{n}$ in $\mathbf{K}$; if -1 is not a sum of squares in $\mathbf{K}$ one puts $s(\mathbf{K}) = \infty$. (Examples: $s(\mathbb{Q}) = \infty = s(\mathbb{R})$, but $s(\mathbb{C}) = 1$. For some further examples see Theorem 5 below!) If $\mathbf{K}$ is a quadratic number field (this means $[\mathbf{K} : \mathbb{Q}] = 2$), then $k = \mathbb{Q}(\sqrt{d})$ for some squarefree integer d. If $d > 0$, clearly $s(\mathbf{K}) = \infty$, and for $d < 0$ $s(\mathbf{K})$ is computed as follows [**FGS**]:

THEOREM 5. *Let m be a squarefree positive integer, then*

$$s\big(\mathbb{Q}(\sqrt{-m})\big) = \begin{cases} 1 & \text{if } m = 1, \\ 2 & \text{if } m \neq 1, \quad m \not\equiv 7 \pmod 8, \\ 4 & \text{if } m \equiv 7 \pmod 8. \end{cases}$$

(An early result closely related to Theorem 5 is [**H**], Theorem 13.)

We are going to show that Theorem 1 and Theorem 5 are *equivalent*: either may be proved from the other. We pave the way by proving three easy lemmas:

LEMMA 6. *For any $a, b, c, d, x, y, z, w \in \mathbb{Q}$ we have*

$$(a^2 + b^2 + c^2 + d^2)(x^2 + y^2 + z^2 + w^2) = (ax + by + cz + dw)^2 + \boxed{3}.$$

Proof. Let H (for Hamilton) denote the ordinary quaternions over $\mathbb{Q}$:

$$H = \mathbb{Q} \cdot 1 \oplus \mathbb{Q} \cdot i \oplus \mathbb{Q} \cdot j \oplus \mathbb{Q} \cdot k$$

with $i^2 = j^2 = -1$, $ij = -ji = k$. For $\alpha = a + bi + cj + dk \in H$ call a the *scalar part* of α; define $\bar{\alpha} = a - bi - cj - dk$ and

$$N(\alpha) = \alpha\bar{\alpha} = a^2 + b^2 + c^2 + d^2 = \boxed{4} \in \mathbb{Q};$$

and observe that $N(\alpha\beta) = N(\alpha)N(\beta)$ for all $\alpha, \beta \in H$. Now to prove the lemma, given $a, b, c, d, x, y, z, w \in \mathbb{Q}$ put $\alpha = a + bi + cj + dk$ and $\beta = x - yi - zj - wk$. Then

$$(a^2 + b^2 + c^2 + d^2)(x^2 + y^2 + z^2 + w^2) = N(\alpha)N(\beta) = N(\alpha\beta).$$

But the scalar part of $\alpha\beta$ is $ax + by + cz + dw$, so $N(\alpha\beta) = (ax + by + cz + dw)^2 + \boxed{3}$, as claimed.

Notice that Lemma 6 has nothing to do with the rational field $\mathbb{Q}$: the statement and proof go through verbatim with any field—indeed, any commutative ring—in place of $\mathbb{Q}$.

LEMMA 7. *Let m be a squarefree positive integer. Then $m = \boxed{3}$ in $\mathbb{Z} \Leftrightarrow m = \boxed{3}$ in $\mathbb{Q} \Leftrightarrow -1 = \boxed{3}$ in $\mathbb{Q}(\sqrt{-m})$.*

Proof. The first "$\Leftrightarrow$" just re-states Lemma 4, so we concentrate on the second. If $-1 = \boxed{3}$ in $\mathbb{Q}$ $(\sqrt{-m})$, then $0 = \boxed{4}$ non-trivially, say $0 = \Sigma(a_i + b_i\sqrt{-m})^2$, where $a_i, b_i, \in \mathbb{Q}$ and Σ means $\Sigma_{i=1}^4$. Thus $\Sigma b_i^2 \neq 0$ (otherwise $b_i = 0$ for all i, a contradiction) and

$$\Sigma a_i^2 - m\Sigma b_i^2 = 0 = \Sigma a_i b_i.$$

Now from $m\Sigma b_i^2 = \Sigma a_i^2$ and Lemma 6 we get

$$m\left(\Sigma b_i^2\right)^2 = \left(\Sigma a_i^2\right)\left(\Sigma b_i^2\right) = \left(\Sigma a_i b_i\right)^2 + \boxed{3}$$

and since $\Sigma a_i b_i = 0$, this gives $m(\Sigma b_i^2)^2 = \boxed{3}$. Dividing by the non-zero square $(\Sigma b_i^2)^2$ shows $m = \boxed{3}$ in $\mathbb{Q}$.

Conversely, if $m = \boxed{3}$ in $\mathbb{Q} \subseteq \mathbb{Q}(\sqrt{-m})$, add the fact that $-m$ is a square to get $0 = \boxed{4}$ non-trivially, hence $-1 = \boxed{3}$ in $\mathbb{Q}(\sqrt{-m})$.

Before giving the last of the three promised lemmas, notice that we already have Theorem 5 $\Rightarrow$ Theorem 1. Indeed, Lemma 7, with Theorem 5, shows that a squarefree positive integer m is $\boxed{3}$ in $\mathbb{Z} \Leftrightarrow m \not\equiv 7 \pmod 8$. This is precisely Corollary 2, which we have already seen is equivalent to Theorem 1.

For the implication Theorem 1 $\Rightarrow$ Theorem 5 we need:

LEMMA 8. *Let m be a positive squarefree integer; then*

(a) $s(\mathbb{Q}(\sqrt{-m})) \leqslant 4$.

(b) *If* $s(\mathbb{Q}(\sqrt{-m})) \leqslant 3$, *then in fact* $s(\mathbb{Q}(\sqrt{-m}) \leqslant 2$; *indeed* $s(\mathbf{K}) \leqslant 3 \Rightarrow s(\mathbf{K}) \leqslant 2$ *for any field* $\mathbf{K}$.

(c) $s(\mathbb{Q}(\sqrt{-m})) = 1 \Leftrightarrow m = 1$.

Proof. Of course, Lemma 8 follows from Theorem 5! However, we want to use it to prove Theorem 5 from Theorem 1, so we need to prove Lemma 8 directly. Our proof *will* use Theorem 1, or rather its consequence Corollary 3. But since our application of the lemma is to show Theorem 1 $\Rightarrow$ Theorem 5, this is not circular, and we may proceed as follows:

(a) From Corollary 3, $m = \boxed{4}$ in $\mathbb{Z} \subseteq \mathbb{Q}(\sqrt{-m})$, and adding the fact that $-m$ is a square gives $0 = \boxed{5}$ nontrivially in $\mathbb{Q}(\sqrt{-m})$, whence $-1 = \boxed{4}$.

For (b), we give first an explicit unenlightening proof: we are given $-1 = a^2 + b^2 + c^2$ in $\mathbf{K}$ and we may assume $1 + a^2 \neq 0$ (otherwise -1 is a square and there is nothing to prove). Then direct calculation verifies

$$-1 = \left(\frac{b - ac}{1 + a^2}\right)^2 + \left(\frac{c + ab}{1 + a^2}\right)^2.$$

A less explicit, though more enlightening, proof of (b) follows: again we start with $-1 = a^2 + b^2 + c^2$, $1 + a^2 \neq 0$. Then, dividing $0 = 1 + a^2 + b^2 + c^2$ by $1 + a^2$, we have $0 = 1 + \dfrac{b^2 + c^2}{1 + a^2}$. Thus

$$-1 = \frac{b^2 + c^2}{1 + a^2} = \frac{(b^2 + c^2)(1 + a^2)}{(1 + a^2)^2} = \frac{\boxed{2} \cdot \boxed{2}}{\boxed{1}} = \boxed{2}.$$

It is worth noting that (b) generalizes: if $-1 = \boxed{n}$ in K and t is chosen so that $2^t \leqslant n < 2^{t+1}$, then $-1 = \boxed{2^t}$ in K. In particular the level of a field, if finite, is a power of two. This generalization fails for rings in general, however. For example -1 is a sum of three squares, but not a sum of two, in the ring of integers modulo 12.

The implication "$\Leftarrow$" in (c) is trivial, and "$\Rightarrow$" is nearly so: $-1 = (a + b\sqrt{-m})^2$ entails $a^2 - mb^2 = -1$ and $ab = 0$; but $b \neq 0$ since $-1 \neq \boxed{1}$ in $\mathbb{Q}$; hence $a = 0$, and then $mb^2 = 1$ forces $m = 1$ since m is squarefree.

This completes the proof of Lemma 8, and we now use it to show that Theorem 5 follows from Theorem 1 (Gauss' Theorem), or from its special case Corollary 2. From Lemma 8 we know that $s(\mathbb{Q}(\sqrt{-m})) = 1$ for $m = 1$ and $S(\mathbb{Q}(\sqrt{-m})) = 2$ or 4 for all $m > 1$. With Lemma 7, Corollary 2 gives $s(\mathbb{Q}(\sqrt{-m})) = 4 \Leftrightarrow m \equiv 7 \pmod 8$, and this secures Theorem 5.

The proof of the equivalence of Theorem 1 and Theorem 5 is now complete, and one may ask for the moral of the story. The interpretation in [**R**] is that Theorem 5, which first appears (as far as I know) as Theorem 7 in [**FGS**], gives a new proof of Gauss' Theorem. Alternatively, one can start with Gauss and view the equivalence as providing a very simple demonstration of Theorem 5. From a slightly loftier (?) point of view, one might use the equivalence to put Theorem 5 in perspective: Gauss' Theorem, however one approaches it, is a substantial result, and the same may therefore be said of Theorem 5 which computes the levels of quadratic number fields.

References

[**FGS**]. B. Fein, B. Gordon, and J. H. Smith, On the representation of -1 as a sum of two squares in an algebraic number field, J. Number Theory, 3 (1971) 310–315.

[**G**]. G. R. Greenfield, Sums of three and four integer squares, Rocky Mountain J. Math., 13 (1983) 169–175.

[**H**]. H. Hasse, Darstellbarkeit von Zahlen durch quadratische Formen in einem beliebigen algebraischen Zahlkörper, Crelle's J., 153 (1923) 113–130.

[**R**]. L. J. Risman, A new proof of the three squares theorem, J. Number Theory, 6 (1974) 282–283.

[**S**]. J.-P. Serre, A Course in Arithmetic, Springer-Verlag, New York, 1973.

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

ANY ANSWERS ANENT THESE ANALYTICAL ENIGMAS?

RICHARD K. GUY

Several unsolved problems have been submitted recently, which are brief to state. Neither the referees nor your editor were able to find relevant references. Here are four examples from analysis.

Alan A. Grometstein, Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA 02173-0073, asks about the function

$$f(x, y) = yx^y\{y^x - (y-1)^x\} - xy^x\{x^y - (x-1)^y\}.$$

He notes that it is antisymmetric: $f(x, y) = -f(y, x)$ and $f(x, x) = 0$.

If x, y are integers, $x > y \geqslant 1$, is $f(x, y) > 0$?

He says that there is a good deal of numerical evidence. The restriction to integers may be inessential.

Louis Funar, Department of Mathematics, University of Craiova, Craiova, A.I. Cuza nr. 13, 1100 Romania, asks if, given an arbitrary function $f\colon \mathbb{R} \to \mathbb{R}$,

Do there exist functions $g\colon \mathbb{R} \to \mathbb{R}$ and $h\colon \mathbb{R} \to \mathbb{R}$,

the first one bijective and the second one injective, such that $f = g + h$?

Bogusław Tomaszewski, Department of Mathematics, Oklahoma State University, OK 74078, considers n real numbers $a_1, \dots, a_n$ such that $\sum_{i=1}^n a_i^2 = 1$. Of the 2^n expressions $|\varepsilon_1 a_1 + \cdots + \varepsilon_n a_n|$ with $\varepsilon_i = \pm 1, 1 \leqslant i \leqslant n$,

Can there be more with value > 1 than with value $\leqslant 1$?

Carl Ponder, Computer Science Division, University of California, Berkeley, CA 94720, defines $\varphi_h(x)$ by the differential equation

$$\frac{d}{dx}\varphi_h(x) = \{\varphi_{h-1}(x)\}^2$$

with boundary conditions $\varphi_0(x) = \varphi_h(0) = 1$, and asks

What is the asymptotic behavior of $\varphi_h(1)$ as $h \to \infty$?

It is not hard to see that $\varphi_h(x)$ is a polynomial in x of degree $2^h - 1$. For example

$$\varphi_1(x) = 1 + x,$$

$$\varphi_2(x) = 1 + x + x^2 + \frac{1}{3}x^3,$$

$$\varphi_3(x) = 1 + x + x^2 + x^3 + \frac{2}{3}x^4 + \frac{1}{3}x^5 + \frac{1}{9}x^6 + \frac{1}{63}x^7.$$

If $\varphi_h(x) = \sum_{i=0}^{2^h-1} a_{h,i}x^i$, then, for $h \geqslant 0$, $a_{h,0} = a_{h,1} = \cdots = a_{h,h} = 1$, and $a_{h,h+1} = a_{h+1,h+3} = 1 - 2^h/(h+1)!$ However, the formulas

$$a_{h,h+3} = 1 - \frac{2^{h-3}}{(h-1)!} - \frac{2^h}{3(h!)} + \frac{5 \cdot 2^h}{(h+3)!} \quad (h \geqslant 2)$$

$$a_{h,h+4} = 1 - \frac{2^{h-4}}{3((h-2)!)} - \frac{2^{h-1}}{3((h-1)!)} - \frac{2^{h-2}}{h!} + \frac{5 \cdot 2^{h-1}}{(h+2)!} + \frac{5 \cdot 2^{h+1}}{(h+4)!} \quad (h \geqslant 3)$$

$$a_{h,h+5} = 1 - \frac{2^{h-7}}{3((h-3)!)} - \frac{2^{h-3}}{3((h-2)!)} - \frac{13 \cdot 2^{h-3}}{9((h-1)!)} - \frac{2^h}{5(h!)}$$

$$+ \frac{5 \cdot 2^{h-3}}{(h+1)!} + \frac{5 \cdot 2^h}{3((h+2)!)} + \frac{5 \cdot 2^h}{(h+3)!} + \frac{119 \cdot 2^h}{(h+5)!} \quad (h \geqslant 4)$$

are increasingly disappointing, though we can say that

$$a_{h,2^h-1} = 1/(2-1)^{2^{h-1}}(2^2-1)^{2^{h-2}}(2^3-1)^{2^{h-3}}\cdots(2^{h-2}-1)^{2^2}(2^{h-1}-1)^2(2^h-1).$$

It is also not hard to show that $0 \leqslant a_{h,i} \leqslant 1$, so we immediately have the bounds

$$h+1 \leqslant \varphi_h(1) \leqslant 2^h.$$

Calculation of the first few values:

$h =$	0	1	2	3	4	5	6	7	8
$\varphi_h(1) =$	1	2	$3\frac{1}{3}$	$5\frac{8}{63}$	7.533	10.747	15.019	20.674	28.131

suggests that, for $h \geqslant 2$, $\varphi_h(1) > (3/2)^h$, but your editor was unable to obtain a convincing proof.

NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.

For instructions about submitting Notes for publication in this department see the inside front cover.

A SIMPLE PROOF OF THE DIRICHLET-JORDAN CONVERGENCE TEST

GEORGES A. LION
Département de Mathématiques, U. E. R. des Sciences de Limoges, 87060 Limoges Cedex, France

Let f be a complex-valued integrable function on the interval $[0, 2\pi]$, and recall that the Fourier coefficients of f are defined by

$$a_k = (1/2\pi)\int_0^{2\pi} f(t)\, e^{-ikt}\, dt,$$

and that the partial sums of the Fourier series of f are defined by

$$S_N f(x) = \sum_{k=-N}^{N} a_k e^{ikx}.$$

The Dirichlet-Jordan convergence test states that if f has bounded variation, then for each x the limit of $S_n f(x)$, as n tends to infinity, exists and is equal to $(f(x^+) + f(x^-))/2$. Here $f(x^+)$ and $f(x^-)$ denote the right and left hand limits of f at x, and as usual we will extend f to be a 2π-periodic function defined on the whole real line.

The original proof of the Dirichlet-Jordan convergence test, as given by Dirichlet for a monotonic function [**1**], and extended by Jordan to functions of bounded variation ([**2**], pp. 264–289), is based upon the second mean value theorem (presented, for example, in [**3**], p. 245). This method is used again by Zygmund ([**4**], pp. 57–58), who also gives another proof using tools from Cesaro summability theory.

The proof presented here is more straightforward and involves slightly less work than the usual proofs. However, the usual proofs have the advantage of giving the extra information that if f is continuous in addition to having bounded variation, then the partial sums $S_n f$ converge uniformly to f, rather than just pointwise.

We begin our proof by recalling that the usual elementary classical computations produce the familiar integral formula for $S_n f(x)$:

$$S_n f(x) = \frac{1}{\pi}\int_0^{\pi/2}\{f(x+2u) + f(x-2u)\}\frac{\sin(2n+1)u}{\sin u}\, du.$$

In particular, we have

$$1=\frac{2}{\pi}\int_0^{\pi/2}\frac{\sin(2n+1)u}{\sin u}\,du.$$

For $u\in[0,\pi/2]$, put $g(u)=f(x+2u)+f(x-2u)$; the function g is of bounded variation, and without changing the integral we may suppose that g is left continuous. So there exists a complex bounded Borel measure μ on $[0,\pi/2]$ such that

$$\mu([0,u))=g(u)-g(0),\quad \text{for } u\in\left[0,\frac{\pi}{2}\right].$$

Then

$$S_nf(x)=\frac{1}{\pi}\int_0^{\pi/2}g(u)\frac{\sin(2n+1)u}{\sin u}\,du=\frac{g(0)}{2}+\frac{1}{\pi}\int_0^{\pi/2}\frac{\sin(2n+1)u}{\sin u}\left(\int_{[0,u)}d\mu(t)\right)du$$

$$=\frac{g(0)}{2}+\frac{1}{\pi}\int_{[0,\pi/2)}\left(\int_t^{\pi/2}\frac{\sin(2n+1)u}{\sin u}\,du\right)d\mu(t),\quad \text{(by Fubini)}.$$

For $t>0$, the integrand in the outside integral on the right tends to 0 pointwise as $n\to\infty$, and remains uniformly bounded for $t\geqslant 0$. The first property is immediate from the Riemann-Lebesgue lemma (or a simple integration by parts), and the second one is a standard property of the Dirichlet kernel $[\sin(2n+1)u/\sin u]$.

By the Lebesgue dominated convergence theorem, we have finally

$$\lim_{n\to\infty}S_nf(x)=\frac{g(0)}{2}+\frac{\mu(\{0\})}{2}=\frac{1}{2}\{f(x^+)+f(x^-)\},$$

completing the proof.

References

1. P. G. Lejeune-Dirichlet, Sur la convergence des séries trigonométriques qui servent à représenter une fonction arbitraire entre des limites données, J. Reine Angew. Math., 4 (1829) 157–169.
2. C. Jordan, Cours d'analyse de l'Ecole Polytechnique, t.2, calcul intégral, Gauthier-Villars, Paris, 1894.
3. B. S. Nagy, Introduction to Real Functions and Orthogonal Expansions, Oxford University Press, New York, 1965.
4. A. Zygmund, Trigonometric Series, vol. 1, Cambridge University Press, 1968.

THE FIXED DIVISOR OF A POLYNOMIAL

JAN TURK
Korianderstraat 72, 7322 NE Apeldoorn, The Netherlands

1. Introduction. Let $F\in\mathbb{Z}[X]$ be not identically zero. The maximal positive integer d with the property that d is a divisor of $F(x)$ for every $x\in\mathbb{Z}$ is called the fixed divisor of F; we denote it by $d(F)$. The maximal positive integer c that divides all coefficients of F is called the content of F; we denote it by $c(F)$. Clearly, $c(F)$ divides $d(F)$. These concepts are well known; see, e.g., [**1**] and [**2**]. We mention some simple examples:

$$F=(X+1)\cdots(X+d)\text{ has } d(F)=d!\quad\text{and}\quad c(F)=1;$$

$$F=X^2+X+2\text{ has } d(F)=2\quad\text{and}\quad c(F)=1.$$

In this note we answer the question whether it is a very unusual property for an $F\in\mathbb{Z}[X]$ to have a fixed divisor $d(F)$ greater than one. It is known that, asymptotically, only 0% of the $F\in\mathbb{Z}[X]$ have $c(F)>1$. However, for the property $d(F)>1$ the percentage is about 28. See Sections 2 and 3 for the exact meaning of these statements. We can actually answer these questions for polynomials in any number of variables. Our arguments involve only elementary

linear algebra and number theory.

2. Definitions. If $F = \sum_{j=0}^{d} f_j X^j \in \mathbb{Z}[X]$ with $f_d \neq 0$, then the degree $\deg(F)$ of F is d and the height $H(F)$ is $\max_{0 \leqslant j \leqslant d} |f_j|$. The fixed divisor $d(F)$ is $\max\{d \in \mathbb{N} | F(\mathbb{Z}) \subset d\mathbb{Z}\}$ and the content $c(F)$ is $\max\{c \in \mathbb{N} | F \in c\mathbb{Z}[X]\}$. For $F = 0$ we define $\deg(F) = -\infty$, $d(F) = \infty$ and $c(F) = \infty$. Note that for $d \in \mathbb{N}_0$ and $H \in \mathbb{N}$, the set $\{F \in \mathbb{Z}[X] | \deg(F) \leqslant d,\ H(F) \leqslant H\}$ is finite. We denote the number of elements of a finite set S by $\#(S)$. For any subset T of $\mathbb{Z}[X]$ one defines the probability that an $F \in \mathbb{Z}[X]$ of degree $\leqslant d$ belongs to T as

$$\operatorname{Prob}(F \in T | \deg(F) \leqslant d) = \lim_{H \to \infty} \frac{\#\{F \in T | \deg(F) \leqslant d,\ H(F) \leqslant H\}}{\#\{F \in \mathbb{Z}[X] | \deg(F) \leqslant d,\ H(F) \leqslant H\}},$$

provided the limit exists. The probability that $F \in \mathbb{Z}[X]$ belongs to T is defined as

$$\operatorname{Prob}(F \in T | F \in \mathbb{Z}[X]) = \lim_{d \to \infty} \operatorname{Prob}(F \in T | \deg(F) \leqslant d),$$

provided this limit exists. We give some examples of this notion of probability.

(1) $\operatorname{Prob}(\deg(F) \leqslant d | F \in \mathbb{Z}[X]) = 0$ for all $d \in \mathbb{N}_0$, trivially,

(2) $\operatorname{Prob}(c(F) = 1 | \deg(F) \leqslant d) = \prod_p (1 - p^{-(d+1)})$,

where the product is over the prime numbers, as is well known and easy to prove; in particular,

$$\operatorname{Prob}(c(F) = 1 | F \in \mathbb{Z}[X]) = 1.$$

(3) $\operatorname{Prob}(\operatorname{Gal}(F) = S_d | \deg(F) \leqslant d) = 1$ for all $d \in \mathbb{N}$, where $\operatorname{Gal}(F)$ is the Galois group of F and S_d the full symmetric group on d symbols, see [**3**, p. 204].

3. Results. The natural basis for the $\mathbb{Q}$-vectorspace $\mathbb{Q}[X]$ is $\{X^j | j \in \mathbb{N}_0\}$. For our purposes it is more appropriate to use the basis $\left\{\binom{X}{j} \middle| j \in \mathbb{N}_0\right\}$, where

$$\binom{X}{j} = (j!)^{-1} \prod_{i=0}^{j-1} (X - i).$$

LEMMA 1. *Let* $F \in \mathbb{Q}[X]$, $F = \sum_{j=0}^{d} F_j \binom{X}{j}$ *and let* $c, D \in \mathbb{Q}$. *Then*

(1) $F \in c\mathbb{Z}[X]$ *if and only if* $F_j \in cj!\mathbb{Z}$ *for* $0 \leqslant j \leqslant d$,

(2) $F(\mathbb{Z}) \subset D\mathbb{Z}$ *if and only if* $F_j \in D\mathbb{Z}$ *for* $0 \leqslant j \leqslant d$.

Proof. Write $F = \sum_{j=0}^{d} f_j X^j$. Compare the coefficients of X^j in the identity

$$\sum_{j=0}^{d} f_j X^j = \sum_{j=0}^{d} (j!)^{-1} F_j \prod_{i=0}^{j-1} (X - i).$$

This gives

$$f_j = (j!)^{-1} F_j + \sum_{k>j}^{d} (k!)^{-1} F_k c_k(j) \quad \text{for } 0 \leqslant j \leqslant d, \quad \text{with } c_k(j) \in \mathbb{Z}. \tag{3}$$

Assume that $F \in c\mathbb{Z}[X]$, i.e., $f_j \in c\mathbb{Z}$ for $0 \leqslant j \leqslant d$. Suppose we have proved that $(k!)^{-1} F_k \in c\mathbb{Z}$ for $j < k \leqslant d$ and some $0 < j < d$. Then if follows from (3) and $f_j \in c\mathbb{Z}$ that $(j!)^{-1} F_j \in c\mathbb{Z}$. Since, by (3) with $j = d$, $(d!)^{-1} F_d \in c\mathbb{Z}$, it follows successively that $(k!)^{-1} F_k \in c\mathbb{Z}$ for $k = d$, $d - 1, \ldots, 0$. The other implication in (1) follows immediately from (3).

To prove (2) we observe that

$$F(j) = F_j + \sum_{k=0}^{j-1} F_k \binom{j}{k} \quad \text{for } 0 \leqslant j \leqslant d. \tag{4}$$

Assume that $F(\mathbb{Z}) \subset D\mathbb{Z}$, hence $F(j) \in D\mathbb{Z}$ for $0 \leqslant j \leqslant d$. Suppose we have proved that $F_k \in D\mathbb{Z}$ for $0 \leqslant k < j$ for some $0 < j < d$. Then it follows from (4) and $F(j) \in D\mathbb{Z}$ that $F_j \in D\mathbb{Z}$. Since, by (4) with $j = 0$, $F_0 = F(0) \in D\mathbb{Z}$, it follows successively that $F_k \in D\mathbb{Z}$ for $k = 0, 1, \ldots, d$. The other implication in (2) follows immediately from $F(x) = \sum_{k=0}^{d} F_k \binom{x}{k}$ for $x \in \mathbb{Z}$.

LEMMA 2. *Let $D \in \mathbb{N}$ and $d \in \mathbb{N}_0$. Then*

$$\operatorname{Prob}(F(\mathbb{Z}) \subset D\mathbb{Z} \mid F \in \mathbb{Z}[X], \deg(F) \leqslant d) = \prod_{i=0}^{d} \left(D^{-1}\gcd(i!, D)\right).$$

Proof. Let $H \in \mathbb{N}$. We must calculate the fraction N_H/D_H and take the limit for $H \to \infty$, where

$$N_H = \#\{F \in \mathbb{Z}[X] \mid \deg(F) \leqslant d,\ H(F) \leqslant H,\ F(\mathbb{Z}) \subset D\mathbb{Z}\}$$

and

$$D_H = \#\{F \in \mathbb{Z}[X] \mid \deg(F) \leqslant d,\ H(F) \leqslant H\}.$$

Clearly, $D_H = (2H+1)^{d+1}$.

The condition $F(\mathbb{Z}) \subset D\mathbb{Z}$ can be conveniently expressed in terms of the F_j; from Lemma 1 we obtain

$$N_H = \#\left\{\left(F_j\right)_{j=0}^{d} \mid F_j \in j!\mathbb{Z} \cap D\mathbb{Z},\ |f_j| \leqslant H \quad \text{for } 0 \leqslant j \leqslant d\right\}.$$

As $j!\mathbb{Z} \cap D\mathbb{Z} = \operatorname{lcm}(j!, D)\mathbb{Z}$ we put $x_j = F_j/\operatorname{lcm}(j!, D)$ for $0 \leqslant j \leqslant d$ to obtain

$$N_H = \#\left\{\left(x_j\right)_{j=0}^{d} \mid x_j \in \mathbb{Z},\ |f_j| \leqslant H \quad \text{for } 0 \leqslant j \leqslant d\right\}.$$

In order to express the conditions $|f_j| \leqslant H$ in terms of the x_j we rewrite (3) as follows.

$$f_j = \sum_{k=j}^{d} c_k(j) F_k/k! = \sum_{k=j}^{d} \left(c_k(j)\operatorname{lcm}(k!, D)/k!\right) x_k \quad \text{for } 0 \leqslant j \leqslant d. \tag{5}$$

In other words,

$$\left(f_j\right)_{j=0}^{d} = M\left(x_k\right)_{k=0}^{d}, \tag{6}$$

where $M = (m_{jk})_{j,k=0}^{d}$ is the triangular matrix with entries

$$m_{jk} = c_k(j)\operatorname{lcm}(k!, D)/k! = c_k(j)\, D/\gcd(k!, D).$$

Note that

$$\det(M) = \prod_{k=0}^{d} m_{kk} = \prod_{k=0}^{d} \left(D/\gcd(k!, D)\right).$$

The transformation (6) maps the rectangular box

$$B_H = \left\{\left(f_j\right)_{j=0}^{d} \in \mathbb{R}^{d+1} \mid |f_j| \leqslant H \quad \text{for } 0 \leqslant j \leqslant d\right\}$$

onto the parallelepiped

$$M^{-1}(B_H) = \left\{\left(x_k\right)_{k=0}^{d} \in \mathbb{R}^{d+1} \mid M\left(x_k\right)_{k=0}^{d} \in B_H\right\}.$$

We have shown that

$$N_H = \#\left\{\left(x_k\right)_{k=0}^{d} \in M^{-1}(B_H) \mid x_k \in \mathbb{Z} \quad \text{for } 0 \leqslant k \leqslant d\right\},$$

i.e., N_H equals the number of lattice points in the parallelepiped $M^{-1}(B_H)$. So N_H can be

approximated by the volume of $M^{-1}(B_H)$, with an error $E_d(H)$ which is at most CH^d in absolute value, where $C = C(d)$ does not depend on H. Since the volume of $M^{-1}(B_H)$ equals $\det(M^{-1})$ times the volume of B_H we obtain

$$\lim_{H\to\infty}(N_H/D_H) = \lim_{H\to\infty}\left(\det(M^{-1}) + E_d(H)/(2H+1)^{d+1}\right) = \det(M^{-1}).$$

Since $\det(M^{-1}) = (\det(M))^{-1} = \prod_{k=0}^{d}(D^{-1}\gcd(k!, D))$, we have proved Lemma 2.

LEMMA 3. *Let $D \in \mathbb{N}$ and $d \in \mathbb{N}_0$. Then*

$$(7)\qquad \operatorname{Prob}(d(F) = D \mid F \in \mathbb{Z}[X], \deg(F) \leqslant d) = \sum_{n=1}^{\infty}\mu(n)\prod_{i=0}^{d}\left((nD)^{-1}\gcd(i!, nD)\right).$$

In particular,

$$(8)\qquad \operatorname{Prob}(d(F) = 1 \mid F \in \mathbb{Z}[X], \deg(F) \leqslant d) = \prod_{p}\left(1 - p^{-\min(d+1,p)}\right).$$

Proof. Put

$$f(D) = \operatorname{Prob}(F(\mathbb{Z}) \subset D\mathbb{Z} \mid F \in \mathbb{Z}[X], \deg(F) \leqslant d)$$

and

$$g(D) = \operatorname{Prob}(d(F) = D \mid F \in \mathbb{Z}[X], \deg(F) \leqslant d)$$

for $D \in \mathbb{N}$. Then we have

$$f(D) = \sum_{n=1}^{\infty} g(nD) \quad \text{for } D \in \mathbb{N}.$$

Hence, by Möbius inversion,

$$g(D) = \sum_{n=1}^{\infty}\mu(n)f(nD) \quad \text{for } D \in \mathbb{N}.$$

When $D = 1$ we can rewrite the obtained expression as follows, since $n \to f(n)$ is a multiplicative function:

$$g(1) = \sum_{n=1}^{\infty}\mu(n)f(n) = \prod_{p}(1 - f(p)) = \prod_{p}\left(1 - \prod_{i=0}^{d}p^{-1}\gcd(i!, p)\right) = \prod_{p}\left(1 - p^{-\min(d+1,p)}\right).$$

The product is over the primes p.

THEOREM 1.

$$\operatorname{Prob}(d(F) = 1 \mid F \in \mathbb{Z}[X]) = \prod_{p}(1 - p^{-p}).$$

Proof. Let d go to infinity in (8).

Since $\prod_p(1 - p^{-p})$ is about .722, Theorem 1 explains the statement in the introduction that the percentage of the $F \in \mathbb{Z}[X]$ with $d(F) > 1$ is about 28.

4. Several Variables. The method in Section 3 can be generalized to cover integer polynomials in n variables, where $n \in \mathbb{N}$ is arbitrary. Lemma 2 is a special case of

LEMMA 4. *Let $D \in \mathbb{N}$ and $d_1, \ldots, d_n \in \mathbb{N}_0$. Then*

$$\operatorname{Prob}\left(F(\mathbb{Z}^n) \subset D\mathbb{Z} \mid F \in \mathbb{Z}[X_1, \ldots, X_n],\ \deg_{X_i}(F) \leqslant d_i \quad \text{for } 0 \leqslant i \leqslant n\right)$$

$$= \prod_{\substack{(i_1,\ldots,i_n)\\ 0 \leqslant i_t \leqslant d_t}}\left(D^{-1}\gcd\left(D, \prod_{t=1}^{n} i_t!\right)\right).$$

Lemma 3 generalizes, too. For simplicity we only state the generalization for $D = 1$.

LEMMA 5. *Let $n \in \mathbb{N}$ and $d_1, \dots, d_n \in \mathbb{N}_0$. Then*

$$\operatorname{Prob}\big(d(F) = 1 | F \in \mathbb{Z}[X_1, \dots, X_n], \quad \deg_{X_i}(F) \leqslant d_i \quad \textit{for } i \leqslant i \leqslant n\big)$$

$$= \prod_p \left(1 - p^{-\prod_{i=1}^{n} \min(d_i + 1, p)}\right).$$

Taking the limit for $d_i \to \infty$, $1 \leqslant i \leqslant n$, we find

THEOREM 2. *Let $n \in \mathbb{N}$. Then*

$$\operatorname{Prob}\big(d(F) = 1 | F \in \mathbb{Z}[X_1, \dots, X_n]\big) = \prod_p \left(1 - p^{-p^n}\right).$$

Put $\delta_n = \prod_p (1 - p^{-p^n})$. If we consider only the first three decimal digits of the deltas, we have $\delta_1 = .722$, $\delta_2 = .937$, $\delta_3 = .996$, while the δ_n with $n \geqslant 4$ cannot be distinguished from 1 yet. So we can summarize as follows.

5. Summary. Asymptotically, the percentage of the polynomials in one variable with integral coefficients with fixed divisor exceeding 1 is about 28. If the number of variables is two, this percentage is about 6, and for $n \geqslant 3$ variables this percentage is less than 1, but positive.

References

1. T. Nagell, Introduction to Number Theory, Chelsea, New York, 1964.
2. A. Schinzel and W. Sierpiński, Sur certaines hypothèses concernant les nombres premiers, Acta Arith., 4 (1958).
3. B. van der Waerden, Algebra I, Springer-Verlag, Berlin-Heidelberg, New York, 1971.

INTEGER POLYNOMIALS THAT ARE REDUCIBLE MODULO ALL PRIMES

ROLF BRANDL
Mathematisches Institut, Am Hubland 12, D-8700 Würzburg, West Germany

It is well known that a monic integer polynomial is irreducible over the rationals if it is irreducible modulo some prime. Indeed, this is the idea behind some tests for irreducibility. The converse of this, however, is not true in general as first has been observed by D. Hilbert [**4**]. Since then various other irreducible integer polynomials that are reducible modulo all primes have been found. In [**3**] S. W. Golomb showed that for all positive integers k which are not of the form $k = 1, 2, 4$, q^t or $2q^t$, where q is an odd prime, the cyclotomic polynomial $\Phi_k(X)$ has the above property. By some more sophisticated methods it is possible to construct such polynomials f of degree n, where n is even or is not square-free [**5**, p. 139]. Moreover, the degree of f cannot be a prime [**5**, p. 139]. The question of which degrees n actually can occur then arises. The results above neither prove nor disprove the existence of a suitable polynomial f of degree $n = 15$.

A complete answer to this question may be found in the following theorem.

THEOREM. *Let $n \neq 1$ be a positive integer. Then there exists a monic irreducible integer polynomial of degree n that is reducible modulo all primes if and only if n is not a prime.*

The proof of this result depends on the classical Frobenius Density Theorem:

LEMMA 1 ([**2**], [**5**, p. 133ff.]). *Let f be an irreducible polynomial of degree n over the field Q of the rationals, let N be the splitting field of f and let $G = \operatorname{Gal}(N|Q)$ act on the set of the roots of f. If G contains a permutation φ which is the product of disjoint cycles of length $n_1, \dots, n_t$, then there exists*

an infinite set P_φ of primes such that for any $p \in P_\varphi$ we have the following decomposition of f modulo p:

$$f(X) \equiv \prod_{i=1}^{t} f_i(X) \bmod p,$$

where all f_i are irreducible mod p *and the degree of f_i is n_i.*

The other result that we need is due to R. Dedekind. It is contained in a letter to F. G. Frobenius dated June 8, 1882 (see [**2**] or [**7**, p. 203]).

LEMMA 2. *Let the notation be as in Lemma* 1, *and let f be monic. Let p be a prime not dividing the discriminant of N. If $f(X) \equiv \prod_{i=1}^{t} f_i(X)$* mod p, *where the f_i are irreducible* mod p, *then G contains at least one permutation which is the product of t disjoint cycles of lengths $n_1, \ldots, n_t$, where n_i is the degree of f_i.*

Proof of Theorem. If the degree n of f is a prime, then the Galois group G of f acts transitively on the set of the n roots of f, and so G possesses an element of order n which must act cyclically on the roots of f and the theorem follows from Lemma 1 (see also [**5**, p. 139]).

Now let n be composite. To prove the result, assume that we have constructed a transitive soluble permutation group (G, Ω) of degree n which does not possess any cycle of length n. By a celebrated result of I. R. Šafarevič [**6**], there exists a Galois extension $N|Q$ with $\mathrm{Gal}(N|Q) \cong G$. Let U be the stabiliser of some point in Ω, and let F be the fixed field of U. Let ϑ be a primitive element of F over Q which is an algebraic integer. Then G acts transitively on the set Δ consisting of all roots of the minimal polynomial f of ϑ over the rationals. Moreover, the permutation representations of G on Ω and Δ are similar and so there does not exist any n-cycle in (G, Δ). By the result of R. Dedekind cited above, f is reducible modulo all primes that do not divide the discriminant of N. But the remaining primes are ramified and so [**1**] f has a quadratic factor and we are done.

We are now going to construct our group G. First, assume that n is not a power of a prime. Let $n = p^\alpha m$, where m is not divisible by the prime p. Furthermore, let

$$\Omega = \{1, 2, \ldots, n\}$$

and let

$$a_1 = (1, 2, \ldots, p^\alpha), a_2 = (p^\alpha + 1, p^\alpha + 2, \ldots, 2p^\alpha), \ldots, a_m = (p^\alpha(m-1) + 1, \ldots, n)$$

and

$$b = (1, p^\alpha + 1, 2p^\alpha + 1, \ldots, (m-1)p^\alpha + 1)(2, p^\alpha + 2, 2p^\alpha + 2, \ldots) \cdots .$$

It is straightforward to check that $a_i a_j = a_j a_i$ for all i, j, and $b^{-1} a_i b = a_{i+1}$, where the indices are read modulo m. Let A be the group generated by $a_1^{-1}a_2, a_2^{-1}a_3, \ldots, a_m^{-1}a_1$ and let $B = \langle b \rangle$. Then A is abelian and consists of all elements of the form $a_1^{\alpha_1} \cdots a_m^{\alpha_m}$ with

$$\alpha_1 + \alpha_2 + \cdots + \alpha_m \equiv 0 \bmod p^\alpha.$$

Moreover $b^{-1}Ab = A$ and so A is a normal subgroup of $G := AB$ and G/A is cyclic, so G is soluble. Also, G acts transitively on Ω as can easily be verified.

Suppose that G contains an n-cycle x, so the order of x equals n. Let $x = ay = ya$ where the order of a is a power of p and p does not divide the order of y. Replacing x by a suitable power, we may assume that $y = b$ and so $x = ab$ has order n. It follows that $a \in A$, say $a = a_1^{\alpha_1} \cdots a_m^{\alpha_m}$. As a and b commute, we have

$$a = b^{-1}ab = a_2^{\alpha_1} a_3^{\alpha_2} \cdots a_1^{\alpha_m}.$$

This implies $\alpha_1 \equiv \alpha_2 \equiv \cdots \equiv \alpha_m \bmod p^\alpha$ and so the above yields $m \cdot \alpha_1 \equiv 0 \bmod p^\alpha$. As p does

not divide m, we arrive at

$$\alpha_1 \equiv \alpha_2 \equiv \cdots \equiv \alpha_m \equiv 0 \bmod p^\alpha$$

which is equivalent to $a = 1$. But then the order of $x = b$ equals m, a contradiction.

If n is a prime power, then there exists a noncyclic group G of order n and we may take the regular representation of G on itself. This completes the proof of our theorem.

REMARK. The last argument in the proof above works for all positive integers n such that there exists a noncyclic group of order n. The smallest composite number for which all groups of that order are cyclic is $n = 15$. One can show that the least degree of a splitting field of an irreducible polynomial of degree 15, which is reducible modulo all primes, is 60 and its Galois group is the alternating group A_5.

We finally present an example of an irreducible integer polynomial of degree 15 that is reducible modulo all primes.

EXAMPLE. Let $g(X) = X^5 - X - 1$ and let $\vartheta_1, \dots, \vartheta_5$ be the roots of g. It is easy to see that g has Galois group isomorphic with the symmetric group S_5.

Let $\vartheta = \vartheta_1\vartheta_2 + \vartheta_3\vartheta_4$ and let f be the minimal polynomial of ϑ over the rationals. A straightforward computation shows

$$\begin{aligned} f(X) = {} & X^{15} + 6X^{13} + 7X^{11} - 21X^{10} - 8X^9 - 109X^8 \\ & -17X^7 - 144X^6 - 355X^5 - 48X^4 + 103X^3 + 5X^2 - 56X + 29. \end{aligned}$$

As S_5 does not contain any element of order 15, the argument above shows that f is reducible modulo all primes. But f is a minimal polynomial and so it is irreducible over the rationals.

REMARK. The idea behind the example above is to replace the deep result of Šafarevič [6] by the much more elementary fact that all symmetric groups are Galois groups over the rationals. However, this does not always work as no symmetric group has a transitive permutation representation of degree 9 not containing any cycle of length 9.

References

1. R. Dedekind, Über die Diskriminanten endlicher Körper, Abh. Kgl. Gesellschaft der Wissenschaften zu Göttingen 29 (1882) 1–56 = R. Dedekind, Gesammelte Abhandlungen Vol. 1, Vieweg, Braunschweig, 1930, 351–396.

2. F. G. Frobenius, Über Beziehungen zwischen den Primidealen eines algebraischen Körpers und den Substitutionen seiner Gruppe, Sber. Preuß. Akad. Berlin (1896) 689–703 = F. G. Frobenius, Gesammelte Abhandlungen Vol. II, Springer, Berlin, 1968, 719–733.

3. S. W. Golomb, Cyclotomic polynomials and factorization theorems, this MONTHLY, 85 (1978) 734–737.

4. D. Hilbert, Über diophantische Gleichungen, Nachr. Ges. der Wissenschaften zu Göttingen (1897) 48–54 = D. Hilbert, Gesammelte Abhandlungen Vol. II, Springer, Berlin, 1933, 384–389.

5. G. J. Janusz, Algebraic Number Fields, Academic Press, New York and London, 1973.

6. I. R. Šafarevič, Construction of fields of algebraic numbers with given soluble Galois group, Izvestija Akad. Nauk SSSR, Ser. Mat., 18 (1954) 525–578.

7. B. L. van der Waerden, Algebra I, Springer-Verlag, Berlin, 1966.

COLORING THE PLATONIC SOLIDS

D. H. LEHMER

Department of Mathematics, University of California, Berkeley, CA 94720

1. Introduction. The five regular solids of Plato, the tetrahedron, the cube, the octahedron, the icosahedron and the dodecahedron, have often been used as examples of Graph Theory problems. The problem of coloring these solids is not one of these problems because to find the number of

ways of coloring the dodecahedron requires the use of an electronic computer and accompanying software. In what follows we present a solution which, for completeness, includes all five Platonic solids.

One would naturally think of coloring the faces of a polyhedron so that no two faces with a common edge would be given the same color. However the Platonic solids enjoy a duality in which vertices and faces are interchanged; the cube and octahedron, and the icosahedron and the dodecahedron are dual while the tetrahedron is self-dual. This means that one may consider coloring the vertices of a Platonic solid instead of its faces, being sure that the two endpoints of any edge are colored differently. This makes our problem one of a familiar class of coloring graphs in general (see, for example, Ore [**1**], Chapter 14).

Suppose one has λ different colors available. Our problem then is to find the number of ways that a given Platonic solid can be vertex colored as a function of λ. Graph Theory tells us that this function is a polynomial in λ, called the *chromatic polynomial*. We must determine this polynomial. We also find the factorization of its discriminant into powers of primes.

2. Edge Reduction. If $\lambda = 0$ or 1, then it is clear that no coloring is possible. Hence the desired polynomial has 0 and 1 as roots. The least positive integer c which is not a root is called the *chromatic number* of the Platonic solid, so that at least c colors are needed to do the job. This number is easily found and is given in the following table along with the number of vertices, edges and faces in each Platonic solid.

Solid	c	v	e	f
Tetrahedron	4	4	6	4
Cube	2	8	12	6
Octahedron	3	6	12	8
Icosahedron	4	12	30	20
Dodecahedron	3	20	30	12

Good pictures of the Platonic solids and their planar graphs can be found in Ore [**1**], pp. 8–9, and in Ore [**2**], p. 104.

It is clear that a direct enumeration of the number of ways of coloring a graph G is not a good way to find the chromatic polynomial $f(G)$. Various recursive procedures have been suggested (see, for example, Ore [**1**] and Whitney [**3**]). The method we employ is called edge reduction and it involves the following.

THEOREM 1. *Let G be any graph with at least one edge. Let E be an edge with endpoints v_1 and v_2. Let G' be the result of destroying E (but not v_1 or v_2). Finally let G'' be the graph obtained from G' by letting v_1 coincide with v_2. Then*

$$f(G) = f(G') - f(G''). \tag{1}$$

Proof. Using λ colors we can color G' by coloring G properly and then coloring G so that v_1 and v_2 are colored the same. This latter coloring is the same as coloring G'' properly. That is,

$$f(G') = f(G) + f(G''),$$

which gives (1).

Starting with a given graph G with e edges, we can select an edge E and replace G by G' and G''. The graph G' has $e - 1$ edges and G'' has $\leqslant e - 1$ edges. We repeat the process with G' or G''. After at most e steps we begin to encounter graphs with no edges. A graph consisting of v vertices and no edges obviously has a chromatic polynomial of x^v. This graph makes a contribution to $f(G)$ of either x^v or $-x^v$ depending on its source. Thus it is possible to make an algorithm out of Theorem 1 to find the chromatic polynomial of any graph.

Carrying out this algorithm we find the following polynomials:

Solid	Chromatic Polynomial
Tetrahedron	$x(x-1)(x-2)(x-3)$
Cube	$x(x-1)(x^6-11x^5+55x^4-159x^3+282x^2-290x+133)$
Octahedron	$x(x-1)(x-2)(x^3-9x^2+29x-32)$
Icosahedron	$x(x-1)(x-2)(x-3)(x^8-24x^7+260x^6-1670x^5+6999x^4$ $-19688x^3+36408x^2-40240x+20170)$
Dodecahedron	$x(x-1)(x-2)(x^{17}-27x^{16}+352x^{15}-2950x^{14}+17839x^{13}$ $-82777x^{12}+305866x^{11}-921448x^{10}+2297495x^9$ $-4783425x^8+8347700x^7-12195590x^6+14808795x^5$ $-14713381x^4+11613602x^3-6892084x^2+2751604x-555984)$.

The polynomial for the icosahedron was given by Whitney [3] in 1933 using a different method.

A word may be said about the program for edge reduction. In the first place a graph with v vertices is presented to the computer by its adjacency matrix. This is a $(0,1)$ matrix whose ith row, jth column element is 1 in case the ith vertex is joined to the jth vertex by an edge and is 0 otherwise. Actually, this matrix is stored as a row vector, the ith row seen as a binary integer. In this way the adjacency matrix for the octahedron, namely,

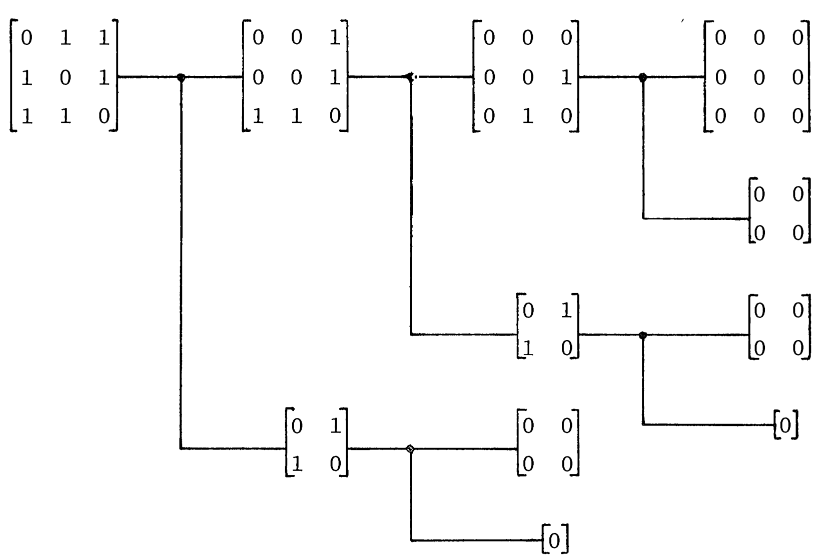
$$\begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

is represented by the vector

$$[29, 46, 51, 51, 29, 46].$$

The various geometrical operations demanded by Theorem 1 that need to be performed on this (or any other) graph are readily transformed into Boolean operations on the elements of this vector.

The edge reduction algorithm may be illustrated by the simple example of the triangle. This gives the following tree of adjacency matrices:



By counting the number of zero matrices of each dimension we get the chromatic polynomial

$$x^3 - 3x^2 + 2x = x(x-1)(x-2),$$

taking into account signs. Since matrices of the same dimension have the same sign, it is evident that the coefficients of any graph with at least one edge are non-zero integers alternating in sign. Also it is evident that the polynomial is monic.

To find the number of ways of coloring each Platonic solid one has only to substitute the number λ of available colors for x in the corresponding chromatic polynomial. The results for $\lambda = 2, 3, 4, 5$ are tabulated below.

	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$
Tetrahedron	0	0	24	120
Cube	2	114	2652	29660
Octahedron	0	6	96	780
Icosahedron	0	0	15600	230400
Dodecahedron	0	7200	168506880	112603286160

3. Discriminants. We recall that the discriminant of a polynomial is defined as the product of the squares of the differences of its roots. The discriminant of the chromatic polynomial of a random graph is of some interest because of its relatively small absolute value. The Platonic solids appear to violate this general rule.

Any chromatic polynomial $f(x)$ can be written

$$f(x) = x^{m_0}(x-1)^{m_1} \cdots (x-k)^{m_k} g(x),$$

where $c = k+1$ is the chromatic number of the graph. Because a graph can be colored with $\lambda + 1$ available colors once it has been colored with λ colors, it follows that the residual polynomial has no rational root. If any of the multiplicities m_i are > 1 the discriminant $D(f)$ is zero. Hence we can restrict ourselves to the case $m_i = 1$.

The following theorem shows that to find the discriminant of f it suffices to find the discriminant of g.

THEOREM 2. *Let f be the chromatic polynomial of the form*

$$f(x) = x(x-1) \cdots (x-k) g(x),$$

where $c = k+1$ is the chromatic number. Then

$$D(f) = \left\{ k(k-1)^2 \cdots 3^{k-2} 2^{k-1} \right\}^2 \left\{ \prod_{i=0}^{k} g(i) \right\}^2 D(g).$$

Proof. By a well-known theorem for the discriminant of a product

$$f(x) = h(x) g(x),$$

$$D(f) = D(h) D(g) [R(h,g)]^2,$$

where

$$R(h,g) = \prod_{h(\alpha_i)=0} g(\alpha_i).$$

Since

$$h(x) = x(x-1) \cdots (x-k),$$

$$D(h) = \left\{ 1^k \cdot 2^{k-1} \cdot (k-1)^2 k \right\}^2,$$

the theorem follows.

The calculation of the discriminant of a monic polynomial of degree 17 (the case for the dodecahedron) using one of today's narrow computers is a matter of some delicacy. This is a case for modular arithmetic. We use the formula (see, for example, Uspensky [4, p. 290])

$$D(g) = \begin{vmatrix} s_0 s_1 & \cdots & s_{16} \\ s_1 s_2 & \cdots & s_{17} \\ \cdots & \cdots & \cdots \\ s_{15} s_{16} & \cdots & s_{31} \\ s_{16} s_{17} & \cdots & s_{32} \end{vmatrix} \tag{2}$$

where $s_j = \sum_{i=1}^{17} \beta_i^j$ and β_i are the roots of $g(x) = 0$.

Let p be any small prime. The coefficients of g can be replaced by their corresponding values modulo p. Then one can apply Newton's formulas for the sums of jth powers of the roots of this polynomial, for $j \leqslant 32$, for substituting into the determinant (2).

By simple row manipulations, using multiplication by the inverse of an element (mod p) to replace division, one can evaluate this determinant modulo p in 17 steps. This was done for the 38 primes

$$p = 2, 3, 5, 7, \ldots, 137, 163, 223, 283, 383.$$

Finally, we applied a simple version of the Chinese Remainder Theorem to determine the exact value for our discriminant. Here we need multiprecision in which multiplication always involves one small factor. In this way we found

$$D(g) = 2^{32} \cdot 23 \cdot 43^2 \cdot 83 \cdot 293 \cdot N,$$

where N is a 43-digit composite number. At this point we called upon Professor Wagstaff who used his EPOC machine to find that

$$N = 72560207 \cdot P,$$

where P is the 35-digit prime given below.

Applying Theorem 2 we find the prime factorization for the discriminants of the Platonic solids to be:

Tetrahedron	$2^4 3^2$
Cube	$-7^2 11^2 13^2 39069367$
Octahedron	$-2^{14} 11^2 59$
Icosahedron	$2^{30} 3^4 5^3 31^2 53 \cdot 67^2 2 7 2^2 821 \cdot 2017^2 \cdot 883529579$
Dodecahedron	$2^{52} 3^{12} 11^2 13^2 43^2 83 \cdot 293 \cdot 1103^2 72560207 \cdot$ 27824298337459013837173524580 3223.

A word may be said about the distribution of the roots of the seventeenth degree polynomial that goes with the dodecahedron. This polynomial has only one real root, namely, 2.66635... . The other eight pairs of complex roots are outside the unit circle and are located in annuli about the origin as follows:

$$\begin{array}{rl} 1 < |z| < 1.5 & \text{four roots,} \\ 2 < |z| < 2.25 & \text{four roots,} \\ 2.25 < |z| < 2.5 & \text{eight roots.} \end{array}$$

References

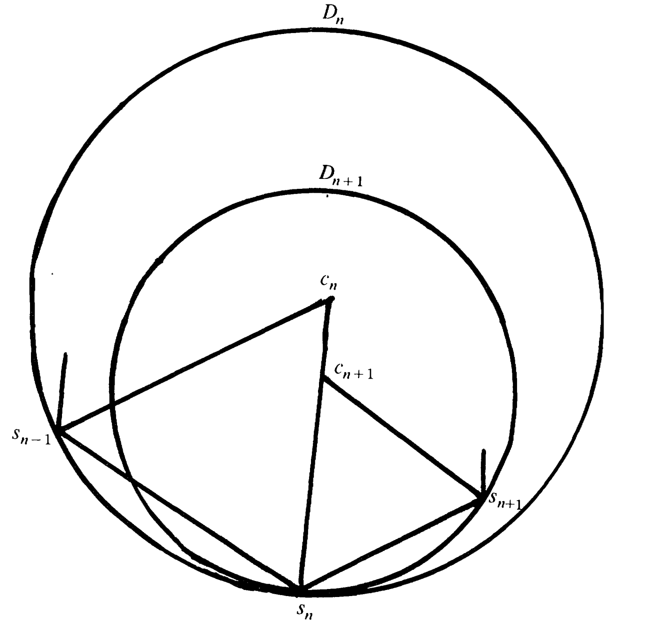
1. Oystein Ore, Theory of Graphs, Amer. Math. Soc. Colloquium Publication V. 38, Providence, 1962.
2. ______, Graphs and Their Uses, Random House, New York, 1963.
3. Hassler Whitney, The coloring of graphs, Ann. of Math., vol. 33, 2 (1932) 688–718.
4. J. V. Uspensky, Theory of Equations, McGraw-Hill, New York, 1948.

THE BROKEN SPIRAL THEOREM

ANATOLE BECK
Department of Mathematics, University of Wisconsin, Madison, WI 53706

One of the classical problems (Abel's Test) set for students in elementary complex analysis is a generalization of the Alternating Series Theorem. It concerns the series $\sum_{n=0}^{\infty} a_n \mathbf{z}^n$, where $\{a_n\}$ is a sequence of positive numbers decreasing to 0 and $|\mathbf{z}| = 1$. The proof that the series converges except possibly at $\mathbf{z} = 1$ is usually done by summation by parts, but this proof does not probe the essence of the result. For many years, I have been using a geometric proof of this theorem, which does not appear in the literature. I offer it here.

I exclude the case $\mathbf{z} = -1$, which is the Alternating Series Theorem. I denote the sum of the first n terms of the series by $\mathbf{s}_n$, and join consecutive values by a line segment, which will have length a_n. Consecutive segments meet at an angle of $\pi - |\arg(\mathbf{z})|$. Let each of these angles be bisected, the bisectors at the ends of the nth segment meeting at a point called $\mathbf{c}_n$, so that the angle $\sphericalangle \mathbf{s}_{n-1}\mathbf{c}_n\mathbf{s}_n = \arg \mathbf{z}$. In each case, the equal sides $\overline{\mathbf{s}_{n-1}\mathbf{c}_n}$ and $\overline{\mathbf{c}_n\mathbf{s}_n}$ are of length $r_n = ta_n$, where $t = \frac{1}{2}\csc|\frac{1}{2}\arg \mathbf{z}|$. Let D_n be the closed disc with center $\mathbf{c}_n$ and radius r_n.



We see that $\mathbf{c}_{n+1}$ lies on the segment $\overline{\mathbf{c}_n\mathbf{s}_n}$, so that $D_{n+1} \subset D_n$, and $\mathbf{s}_{n+1} \in D_{n+1}$. Thus, $\mathbf{s}_k \in D_n$ for all $k \geqslant n$. It is clear that $\bigcap_{n=1}^{\infty} D_n$ consists of a single point, which we denote by $\mathbf{s}$, and that $|\mathbf{s} - \mathbf{s}_n| \leqslant 2r_n = 2ta_n$ for all n. Since $a_n \to 0$, we have $\mathbf{s}_n \to \mathbf{s}$. QED

---

166. MISCELLANEA

What we admire and honor in the masters of our science is the penetrating thought and the broad vision that enabled them to uncover deeply hidden relationships and bring them to the light. The sense of having added an unforeseen truth to the treasury of knowledge is the greatest good fortune and the highest reward to which a mathematician aspires.

—LEON LICHTENSTEIN (1878 -1933), quoted by O. Hölder, Ber. Sächs. Akad. Wiss. 86 (1934) 314.

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

For instructions about submitting material for publication in this department see the inside front cover.

STRONGER VERSIONS OF THE FUNDAMENTAL THEOREM OF CALCULUS

MICHAEL W. BOTSKO AND RICHARD A. GOSSER
Department of Mathematics, Saint Vincent College, Latrobe, PA 15650

In most Advanced Calculus texts the Fundamental Theorem of Calculus is given the following form:

THEOREM. *Let f be Riemann integrable on $[a, b]$ and let g be a function such that $g'(x) = f(x)$ on $[a, b]$. Then*

$$\int_a^b f(x)\,dx = g(b) - g(a).$$

The purpose of this paper is to present two stronger versions of this theorem. In our first result we assume that f is only the right-hand derivative of g. The right-hand derivative of a function g at x is

$$g'_+(x) = \lim_{\Delta x \to 0^+} \frac{g(x + \Delta x) - g(x)}{\Delta x}.$$

THEOREM 1. *Let f be Riemann integrable on $[a, b]$, and let g be a continuous function on $[a, b]$ such that $g'_+(x) = f(x)$ for all x in (a, b). Then*

$$\int_a^b f(x)\,dx = g(b) - g(a).$$

Proof. Let $P = \{x_0, x_1, x_2, \dots, x_n\}$ be any partition of $[a, b]$. Clearly

$$g(b) - g(a) = \sum_k g(x_k) - g(x_{k-1}). \tag{1}$$

Now for each $I_k = [x_{k-1}, x_k]$, by a result in [**2**], there exist c_k and d_k in (x_{k-1}, x_k) such that

$$g'_+(c_k) \leqslant \frac{g(x_k) - g(x_{k-1})}{x_k - x_{k-1}} \leqslant g'_+(d_k). \tag{2}$$

Since $g'_+(x) = f(x)$ for all x in (a, b) we have

$$f(c_k) \leqslant \frac{g(x_k) - g(x_{k-1})}{x_k - x_{k-1}} \leqslant f(d_k).$$

Therefore

$$m_k \leqslant \frac{g(x_k) - g(x_{k-1})}{x_k - x_{k-1}} \leqslant M_k,$$

where $m_k = \inf_{I_k} f(x)$ and $M_k = \sup_{I_k} f(x)$ for each k. Multiplying by $x_k - x_{k-1}$, summing and using (1) we get

$$\sum_k m_k(x_k - x_{k-1}) \leqslant g(b) - g(a) \leqslant \sum_k M_k(x_k - x_{k-1}).$$

Since the partition P was arbitrary we have

$$\underline{\int_a^b} f(x)\,dx \leqslant g(b) - g(a) \leqslant \overline{\int_a^b} f(x)\,dx,$$

where the left and right expressions are the lower and upper Riemann integrals of f respectively. Finally, since f is Riemann integrable, the lower and upper integrals are equal to the Riemann integral and we have our conclusion,

$$\int_a^b f(x)\, dx = g(b) - g(a).$$

As an example let us evaluate $\int_{-1}^{1} f(x)\, dx$ where f is defined as:

$$f(x) = \begin{cases} x^2 + 3 & \text{if } x \geqslant 0, \\ -x & \text{if } x < 0. \end{cases}$$

Clearly f is the derivative of no function on $[-1, 1]$ since it does not have the "intermediate value property". However, if we let g be given by:

$$g(x) = \begin{cases} x^3/3 + 3x & \text{if } x \geqslant 0, \\ -x^2/2 & \text{if } x < 0, \end{cases}$$

it is clear that $g'_+(x) = f(x)$ for all x. Since g is continuous we have by Theorem 1

$$\int_{-1}^{1} f(x)\, dx = g(1) - g(-1) = 23/6.$$

We note that Theorem 1 is already known to be true for the Lebesgue integral, but all proofs that we have seen are much more difficult than the one presented here for the Riemann integral. As an example see [**5**].

Our second stronger form of the Fundamental Theorem of Calculus involves the Schwarz derivative:

$$f^s(x) = \lim_{h \to 0} \frac{f(x+h) - f(x-h)}{2h}.$$

For various interesting theorems on this derivative see [**1**], [**3**], and [**4**].

THEOREM 2. *Let f be Riemann integrable on $[a, b]$, and let g be a continuous function on $[a, b]$ such that $g^s(x) = f(x)$ for all x in (a, b). Then*

$$\int_a^b f(x)\, dx = g(b) - g(a).$$

The proof follows the same pattern as that of Theorem 1 with the Schwarz derivative replacing the right-hand derivative. Inequality (2) is replaced by an analogous inequality involving Schwarz derivatives. (See [**1**, p. 709].)

As an easy example let us evaluate $\int_{-1}^{2} \operatorname{sgn}(x)\, dx$, where

$$\operatorname{sgn}(x) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -1 & \text{if } x < 0. \end{cases}$$

If $g(x) = |x|$, then clearly $g^s(x) = \operatorname{sgn}(x)$ for all x. Thus

$$\int_{-1}^{2} \operatorname{sgn}(x)\, dx = g(2) - g(-1) = 1.$$

We note in closing that the Fundamental Theorem of Calculus follows as an immediate corollary of either Theorem 1 or Theorem 2. Thus in an Advanced Calculus class these two theorems could be established first and the Fundamental Theorem of Calculus then derived from either one of them.

References

1. C. E. Aull, The first symmetric derivative, this MONTHLY, 74 (1967) 708–711.

2. W. J. Knight, Functions with zero right derivatives are constant, this MONTHLY, 87 (1980) 657–658.
3. S. N. Mukhopadhyay, On Schwarz differentiability-V, this MONTHLY, 74 (1967) 542–544.
4. J. Swetits, A note on Schwarz differentiability, this MONTHLY, 75 (1968) 1093–1095.
5. P. L. Walker, On Lebesgue integrable derivatives, this MONTHLY, 84 (1977) 287–288.

WRITING MATHEMATICAL DIALOGUES

GARY BROWN
Department of Mathematics, College of Saint Benedict, Saint Joseph, MN 56374

In January of 1984 I was scheduled to teach a foundations of mathematics course and decided to develop a course which would expose students to controversial issues, historical debates and to the historical development of ideas. The goals for the course were: (1) to expose students to the nature and ideas of mathematics. This included raising such questions as "What is mathematics?", "Why do I enjoy doing mathematics?", "Why study mathematics?"; (2) to show the students through direct experience that studying and doing mathematics can be great fun shared by a group; (3) to give the students an opportunity to use their imaginations and create something from nothing.

I decided that each student in the course would write and act out a play with two other members of the class. The plan was threefold: (1) the opening class discussions would be based upon some of the questions above but structured around the essay by Hammond [**6**] in which an inquirer asks questions to Bers, Sullivan and Puckette; (2) broad introductory readings would be assigned early in the term and (3) groups would meet with me daily for help in developing their plays. More specialized readings could then be assigned.

Students could enroll in this course only with my permission. I interviewed each student to ascertain whether he or she was receptive to writing a group dialogue in mathematics. Most of the students had a major in mathematics and minor in education and were required to take the foundations course for secondary education certification. None had read beyond the normal textbook assignments of his or her past courses.

Now to the details of this plan. The original plan has been partitioned into three parts: (A) "The Early Going"; (B) "Frequent Group Meetings"; (C) "The Presentation of the Plays".

(A) The Early Going. A list of general questions related to ideas from the foundations of mathematics was distributed. These questions fell into one of four categories: foundations of geometry, arithmetic, or analysis, and metamathematics. The questions were stated as simply as possible, yet created so that there were no easy answers. Examples of questions found in the metamathematics category were "Does symbolism enhance or detract from the learning and working of mathematics?", "What are the essential ingredients of any mathematical theory, i.e., what characteristics of a collection of statements must be present for that theory to be called mathematical?"; or from the foundations of analysis category, "Are there different sizes of infinity? For example, does the circle of radius 1 have the same number of points as that of radius 2; does the set of rational numbers have the same size as the set of real numbers?" Questions similar to the above lend themselves to several points of view and hence one can create dramatic tension by writing a conversation involving characters representing different points of view.

A large amount of reading was assigned for the first week of the term. The readings were selected from Davis and Hersch's *Mathematical Experience*, Rucker's *Infinity and the Mind*, and Delong's *Profiles in Mathematical Logic*. These readings gave the beginning student an introduction to mathematical ideas from several authors with different viewpoints.

I gave an introductory lecture on the foundations of mathematics. The content of the lecture focused on the historical development of mathematics and how the subject we now call "foundations" evolved in the 20th century. This was the only lecture on foundations given to the entire class.

I performed a play with a colleague and invited the foundations class together with a math anxiety class to attend the performance. We performed Renyi's play "A Socratic Dialogue on Mathematics" [**4**, pp. 3–25] and afterward led a group discussion. We discussed what major points the author was attempting to make, how he attempted to communicate those points, and how he attempted to make the dialogue entertaining. Finally we discussed the actor's interpretations of the characters and how these interpretations add to or subtract from the intentions of the author.

I lectured one day during the first week about how to write a play. The emphasis was on organization and character development. A teacher of mathematics does not need an extensive theatrical background to discuss these ideas. I found that examining the structure of the play we performed, analyzing our interpretations of the two main characters and gathering advice from colleagues in the theatre arts department were sufficient to write and organize a lecture on playwriting.

Finally, the class was divided into groups of three and each group was asked to write and perform a mathematical play within the next 18 days. This created a certain feeling of urgency and resulted in making all members of the class instantly and continuously active in their goal of performing a play within this short period.

(B) Frequent Group Meetings. After the first week of the term I met with each group at least every other day for the remainder of the semester. These meetings usually required 1 to 1 1/2 hours per group. The groups and individuals developed their plays as follows:

(a) Each group chose a topic. Examples were: (1) "Orwell's 1984 and the Incompleteness Theorem", (2) "Sherlock Holmes sees through Zeno's paradoxes" and (3) "The real numbers and the high school math teacher".

(b) Each participant created a mathematical point of view relevant to the group theme. For example, in the play concerning the use of symbolism in mathematics the participants created two characters, one that modeled Peano's viewpoint that mathematical knowledge is increasingly clarified if we increase the use of symbols in the literature, and the other modeled Hobbes' viewpoint that increased symbolism actually can muddle the simple underlying concepts of mathematics.

(c) Each student wrote a short paper expressing what he or she wished to communicate to the audience. This had to be in the student's own words and was the most difficult aspect of dialogue writing for each student. I was quite critical and often asked them to rewrite their thoughts. Occasionally I was able to push a student to a higher level of mathematics. For example, one student's study of Zeno's paradoxes eventually led him to the discovery of some nontrivial order types. This method of writing a synopsis and being examined orally about it is the best method of evaluation I have encountered for determining the degree of a student's comprehension of a concept.

(d) Each group wrote a general outline of its play that included the name of scenes and subscenes. They were encouraged to enliven their plays by inventing humorous, cute, argumentative or stereotypical characters. We usually met daily for the group to act out a particular scene written the night before. I interjected ideas that might add content or entertainment value. Although criticism was given at each of these meetings, each group was allowed to have complete control of its play. The end result was the creation of four plays each with its own unique style.

(C) The Presentation of the Plays. Each play lasted approximately 45 minutes and was immediately followed by a class discussion. I asked the class after each play what major points the authors were attempting to make, and where the authors succeeded or failed to communicate or entertain the audience. Each group was responsible for typing up a rough draft of the dialogue to be copied and distributed to the class. Each student was required to write a paper synthesizing the concepts considered in the course at the end of the semester.

The writing of mathematical dialogues can be used in many contexts. Ideas for the writing of mathematical dialogues include a play illustrating the definition of the limit of a function, one that compares and contrasts various kinds of algebraic structures, one that compares and contrasts

various kinds of integrals or one that contrasts various inferential tests in statistics. Two excellent sources for further ideas come from Renyi [4] and Lakatos [7]. Renyi's book includes a dialogue on mathematical modeling and Lakatos's book concerns the derivation of Euler's famous formula $V - E + F = 2$.

Teachers of mathematics do not need special knowledge or experience in the writing and presenting of mathematical dialogues. One needs a structured approach (perhaps a list of opening questions) and a willingness to spend enough time early in the process to make this idea work. Once the groups get started the teacher's role is reduced to becoming a good listener. The groups become self-motivated and discover that they can collectively choose a difficult mathematical concept and communicate its essence in a creative fashion.

References

1. P. Davis and R. Hersh, The Mathematical Experience, Houghton Mifflin, Boston, 1981.
2. De Long, Profiles in Mathematical Logic, Addison-Wesley, Reading, MA, 1971.
3. P. Gillette, Calculus and Analytic geometry, 2nd ed., Heath, Lexington, MA, 1984.
4. A. Renyi, Dialogues on Mathematics, Holden-Day, San Francisco, 1967.
5. R. Rucker, Infinity and the Mind, Birkhauser, Boston, 1982.
6. L. A. Steen, Mathematics Today, Twelve Informal Essays, Springer-Verlag, New York, 1978.
7. I. Lakatos, Proofs and Refutations, Cambridge University Press, Cambridge, 1976.

PROBLEMS AND SOLUTIONS

For instructions about submitting ***proposed*** *problems for publication in this department see the inside front cover. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

Solutions *should be sent to the address given on the inside front cover.*

A publishable solution must, above all, be correct. Given corrections, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

For instructions about submitting solutions of these Elementary Problems, which should be mailed by August 31, 1986, see the inside front cover. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgement).

E 3141. *Proposed by Mo Song-Qing, The Beijing 19th Middle School, Beijing, China.*

Suppose $a_i > 0$, $i = 1, 2, \ldots, n$. Consider the inequality

$$\frac{a_1 + a_2 + \cdots + a_k}{k} \cdot \frac{a_2 + a_3 + \cdots + a_{k+1}}{k} \cdot \cdots \cdot \frac{a_n + a_1 + \cdots + a_{k-1}}{k} \geqslant \frac{a_1 + a_2 + \cdots + a_n}{n} (a_1 a_2 \cdots a_n)^{(n-1)/n} \tag{1}$$

for $2 \leqslant k \leqslant n-1$.

(a) Prove or disprove (1) for $n = 3$.
(b)* Prove or disprove (1) for $n > 3$.

E 3142. *Proposed by Zhang Zaiming, Yuxi Teachers' College, China.*

Prove the following refinement of the arithmetic-geometric mean inequality:

$$\sqrt{ab} = G(a,b) \leqslant L(a,b) \leqslant I(a,b) \leqslant A(a,b) = (a+b)/2 \quad \text{for} \quad 0 < a \leqslant b,$$

where the logarithmic mean is

$$L(a,b) = \begin{cases} (b-a)/(\log_e b - \log_e a) & \text{if } 0 < a < b, \\ a & \text{if } 0 < a = b, \end{cases}$$

and the identric mean is

$$I(a,b) = \begin{cases} \frac{1}{e}\left(a^a/b^b\right)^{1/(b-a)} & \text{if } 0 < a < b, \\ a & \text{if } 0 < a = b. \end{cases}$$

(Reference: M. E. Mays, Functions which parameterize means, this MONTHLY, 90 (1983) 677–683.)

E 3143. *Proposed by Allen J. Schwenk, Western Michigan University, Kalamazoo, MI.*

A riffle shuffle of a deck of cards is the commonly used technique of cutting the deck into two portions (not necessarily equal), then, elevating slightly the corners, allowing each portion to fall card by card (not necessarily alternating) merging with the other portion, and finally pushing them together to reconstitute the pack. Given a deck of n cards in arbitrarily permuted order π, determine as a function of π the minimum number of riffle shuffles that could possibly produce the identity sequence $i = 1, 2, \ldots, n$. Describe a procedure that attains this minimum. Which original sequences require the most shuffles?

E 3144. *Proposed by Edwin Buchman, California State University, Fullerton.*

Determine the maximum number of sets in a topological space which can be generated from one set by application (possibly repeated, in any order) of the operations of taking the complement, interior, and boundary of a set.

E 3145. *Proposed by Clinton J. Kolaski, University of Minnesota, Duluth.*

Show that

$$\int_0^{\pi}\left[\frac{\cos nx - \cos ny}{\cos x - \cos y}\right] dx = \pi\frac{\sin ny}{\sin y}; \quad n = 0, 1, 2, \ldots.$$

E 3146*. *Proposed by Walther Janous, Ursulinengymnasium, Innsbruck, Austria.*

Prove or disprove that if a triangle has sides a, b, c, and $2s = a + b + c$, then

$$2s(\sqrt{s-a} + \sqrt{s-b} + \sqrt{s-c}) \leqslant 3\left(\sqrt{bc(s-a)} + \sqrt{ca(s-b)} + \sqrt{ab(s-c)}\right).$$

SOLUTIONS OF ELEMENTARY PROBLEMS

Relatively Prime Variables

E 2953 [1982, 424]. *Proposed by John J. Wahl, Mt. Pocono, PA.*

Let A, B, X, Y be variables subject to the condition $AX - BY = 1$.

(a) Find explicit polynomials u and v in A, B, X, Y with integer coefficients such that

$A^4u - B^4v = 1$.

(b) Prove in fact that for any positive m and n there are u and v such that $A^m u - B^n v = 1$.

Solution by Joseph Baruch Muskat, Bar-Ilan University, Ramat-Gan, Israel. Since

$$AX - BY = 1, \quad (AX - BY)^{n+m-1} = 1 = \sum_{j=0}^{n+m-1} \binom{n+m-1}{j} (AX)^{n+m-1-j}(-BY)^j,$$

by the binomial theorem. The binomial expansion has $n + m$ terms. Each of the first n terms is divisible by A^m, while each of the last m terms is divisible by B^n. Since all the binomial coefficients are integral, we may set

$$u = A^{-m} \sum_{j=0}^{n-1} \binom{n+m-1}{j} (AX)^{n+m-1-j}(-BY)^j,$$

$$v = -B^{-n} \sum_{j=n}^{n+m-1} \binom{n+m-1}{j} (AX)^{n+m-1-j}(-BY)^j.$$

In particular, if $m = n = 4$, then

$$u = A^3X^7 - 7A^2X^6BY + 21AX^5B^2Y^2 - 35X^4B^3Y^3,$$
$$v = B^3Y^7 - 7AXB^2Y^6 + 21A^2X^2BY^5 - 35A^3X^3Y^4.$$

Also solved by M. D. Ašić (Yugoslavia), L. Béla and B. Lippai (Hungary), J. C. Binz (Switzerland), R. Breusch, D. M. Broline, P. S. Bruckman, D. E. Cameron, F. W. Dodd and L. E. Mattics, C. W. Dodge, N. J. Fine, N. Franceschine, V. Grinberg, M. S. Klamkin (Canada), L. Kuipers (Switzerland), L. C. Larson, C. Levesque and G. Lord (Canada), O. P. Lossers (The Netherlands), A. Markovich (student, Yugoslavia), V. D. Mascioni (Switzerland), J. B. M. Melissen (The Netherlands), R. Nelson, R. Pemantle, and P. Y. Wu (Taiwan).

A Unique Type of Partition of the Positive Integers

E 2977 [1982, 756]. *Proposed by Stan Wagon, Smith College.*

Martin Gardner once mentioned the following problem in his *Scientific American* column: Partition the positive integers into two sets, A, B, such that neither $A + A$ or $B + B$ contains a prime. ($X + X$ denotes $\{x + y\colon x, y \in X, x \neq y\}$.) Show that there is a *unique* solution to this problem. (See advanced proposal 6413 [1982, 788; 1984, 372].)

Solution I *by Merrill Barnebey, University of Wisconsin-La Crosse.* The rather obvious way to partition the positive integers into two sets, such that neither $A + A$ nor $B + B$ contains a prime, is to partition them into odd and even integers. If either set of any solution to the given partitioning problem contains two elements of opposite parity, then the Dirichlet theorem insures us we can form a prime, indeed a prime sequence. Thus, the suggested partition is unique.

Solution II *by Lorraine L. Foster, California State University, Northridge.* Inductively assume $1, 3, \ldots, 2k - 1 \in A$ and $2, 4, \ldots, 2k \in B$ for some $k \geqslant 1$. By Bertrand's "Postulate" there exists a prime p such that $2k + 2 < p < 4k + 2$. Easily, $p - (2k + 1) \in B$ and $p - (2k + 2) \in A$ so that $2k + 1 \in A$ and $2k + 2 \in B$ and we are finished.

Also solved by 36 other readers and the proposer. T. E. Moore noted that the problem appeared in the *Journal of Recreational Mathematics*, volume 9, number 4, pages 314–315, where there was also a Fibonacci variant involved.

Strictly Isosceles Spherical Triangles with Three Equal Medians

E 2981 [1983, 54]. *Proposed by Murray S. Klamkin, University of Alberta, Canada.*

If the three medians of a spherical triangle are equal, must the triangle be equilateral? Note

that the sides of a (proper) spherical triangle are *minor* arcs of great circles and thus its perimeter is $< 2\pi$.

Composite solution. Several incorrect solutions were submitted. The following is a composite solution, portions of which were contributed by C. Gorsch, W. Meyer, the proposer, and the editors.

No, surprisingly the triangle need not be equilateral; however, it must be isosceles, and is otherwise severely limited.

Let a, b, c denote the angles subtended at the center of the sphere by the sides of the triangle; let m_a, m_b, m_c likewise denote the angles subtended by the medians from a, b, c respectively.

Using dot products, or the spherical law of cosines, or other means, the following may easily be shown:

$$\cos m_a = \frac{\cos b + \cos c}{2\cos a/2}, \quad \cos m_b = \frac{\cos c + \cos a}{2\cos b/2}, \quad \cos m_c = \frac{\cos a + \cos b}{2\cos c/2}.$$

If the medians are equal—i.e., $m_a = m_b = m_c$— then

$$\frac{\cos b + \cos c}{2\cos a/2} = \frac{\cos c + \cos a}{2\cos b/2} = \frac{\cos a + \cos b}{2\cos c/2},$$

and conversely.

It is clear that these equations hold if $a = b = c$. Moreover, they cannot hold if a, b, c are all different. However, we will show that they may hold if two are equal but the third is different, i.e., the triangle is isosceles but not equilateral.

Suppose, then, that $b = c$. The condition for equality of the medians becomes

$$\frac{\cos a + \cos b}{2\cos c/2} = \frac{\cos b + \cos c}{2\cos a/2}.$$

Let $x = \cos a$ and $y = \cos b = \cos c$. Then, using the half-angle formula, and cancelling common 2's, we obtain

$$\frac{2y}{\sqrt{1+x}} = \frac{x+y}{\sqrt{1+y}}.$$

The graph of this equation may be shown by standard methods of analytic geometry to consist of the line $y = x$ together with a portion of the ellipse $x^2 + 3xy + 4y^2 + x + 3y = 0$. The line may be disregarded completely, as it corresponds to the case of equilateral triangles. The major axis of the ellipse is inclined at $1/4$ of a right angle clockwise from the x-axis. The ellipse is otherwise difficult to describe nicely; but it is easily verified that it contains the following points:

$$A(0,0),\ B(1,-1/2),\ C(9/7,-6/7),\ D(1,-1),\ E(0,-3/4),$$
$$F(-1/2,-1/2),\ G(-1,0),\ \text{and } H(-5/7,1/7).$$

Points C and G are extreme in the x direction; points D and H are extreme in the y direction.

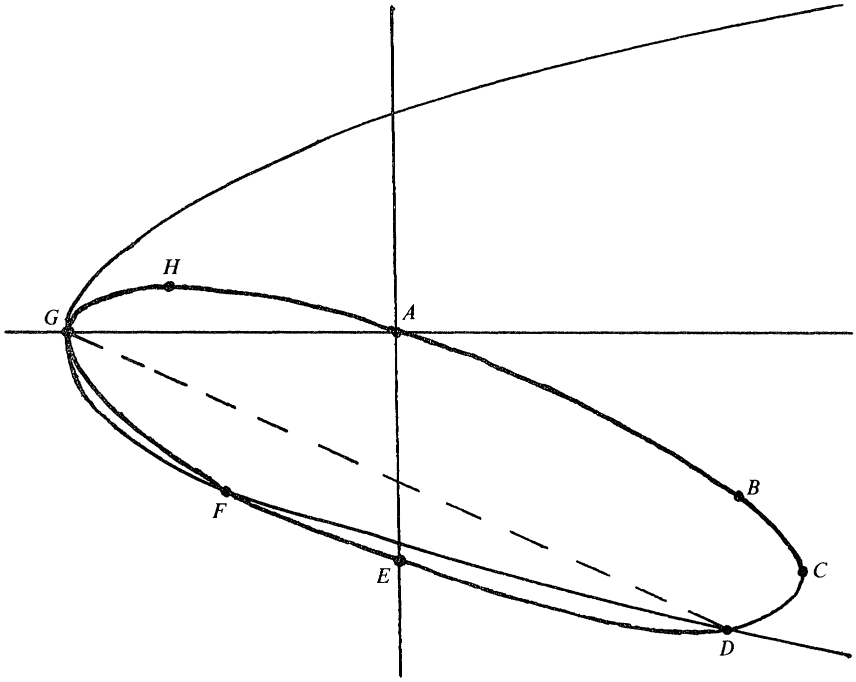
The relevant portion of the ellipse is below the line $zy = -(x+1)$—the rest is introduced as an "extraneous root". This line passes through G and D.

Not even all points (x, y) on this portion of the ellipse will do, however. Some lead to improper spherical triangles. To be proper, the perimeter $(a + 2b)$ cannot exceed 2π. On the other hand, $a \geqslant 2b$ would lead to a flat or impossible triangle. Hence, $a/2 < b < \pi - a/2$. Since the cosine is decreasing over $[0, \pi]$,

$$\cos a/2 > \cos b > \cos(\pi - a/2) = -\cos(a/2).$$

Thus, $\cos^2 b < \cos^2 a/2 = (1 + \cos a)/2$, or $2y^2 < x + 1$.

This corresponds to the region inside a parabola opening along the positive x axis and passing through $(-1,0)$ and $(1, \pm 1)$.



The parabola intersects the ellipse at $G(-1,0)$, $F(-1/2,-1/2)$, and $D(1,-1)$. Point G is a "double root" and, though the parabola touches the ellipse here, it remains outside it. The parabola is inside the ellipse from F to D. Only the portion of the ellipse within the parabola corresponds to viable cases—this is the section from G to F. With this restriction, $a = \cos^{-1}x$ is between 120° and 180°, whereas $b = \cos^{-1}y$ is between 90° and 120°. Any such values correspond to triangles meeting the required condition.

C. Gorsch has noted that though these values lead to proper triangles with the desired property, they are "barely" proper in that their perimeters are all very close to the limiting value 2π. In fact he alleges the existence of a single extremum at approximately $b = 102°$ for which the perimeter is least; but even so the perimeter there is within 4° of 360°.

W. Meyer has noted that the triangle determined by the midpoints of the sides of the original triangle has the property that its two equal sides are 120° each and the remaining side is between 90° and 120°. He asserts that, in fact, an arbitrary triangle meeting these conditions may be given, and then a triangle with the desired medians-equal property may easily be circumscribed around it.

The proposer raises two related questions: whether the triangle must be equilateral if the three altitudes are equal; and likewise if the angle bisectors are equal. He alleges that in the former case, $\sin a = \sin b = \sin c$, and thus the triangle again need not be equilateral due to the rise-fall behavior of sine in $[0, \pi]$. He leaves open the apparently more difficult second question.

How Big Is $\sum_{k=1}^{n} k^{\frac{1}{k}}$?

E 2988 [1983, 134]. *Proposed by Mihaly Bencze, Brasov, Romania.*

For what values of n do the inequalities

$$n + \ln\frac{n+2}{3} < 1 + \sqrt{2} + \sqrt[3]{3} + \cdots + \sqrt[n]{n} < 2n - \ln(n+1)$$

hold?

Solution by William A. Newcomb, Lawrence Livermore National Laboratory, Livermore, CA. The proposed inequalities are correct for all $n \geqslant 2$.

Proof. If $n \geqslant 2$, then

$$n + \ln\frac{n+2}{3} = 1 + \sum_2^n\left(1 + \ln\left(1 + \frac{1}{k+1}\right)\right) < 1 + \sum_2^n\left(1 + \frac{1}{k+1}\right)$$

$$\leqslant 1 + \frac{4}{3} + \sum_3^n\left(1 + \frac{1}{k}\right) \leqslant 1 + \left(\frac{16}{9}\right)^{1/2} + \sum_3^n\left(1 + \frac{\ln k}{k}\right)$$

$$< 1 + 2^{1/2} + \sum_3^n \exp\left(\frac{1}{k}\ln k\right) = \sum_1^n k^{1/k}$$

$$= \sum_1^n (1 + k - 1)^{1/k} \leqslant \sum_1^n\left(\left(1 + \frac{k-1}{k}\right)^k\right)^{1/k} = \sum_1^n\left(2 - \frac{1}{k}\right)$$

$$< \sum_1^n\left(2 - \ln\left(1 + \frac{1}{k}\right)\right) = 2n - \ln(n+1).$$

(The sum from 3 to n, when $n = 2$, should be understood to vanish.)

V. D. Mascioni, M. S. Perkins, O. G. Ruehr, and R. E. Shafer all provided sharper results. Ruehr, in particular, used the Euler-Maclaurin summation formula to obtain the inequalities

$$n + \frac{1}{2}\ln^2 n < \sum_{k=1}^n k^{\frac{1}{k}} < n + \frac{1}{2}\ln^2 n + 1.$$

Also solved by 35 other readers and the proposer.

Tails of Convergent Positive Term Series

E 2996 [1983, 334]. *Proposed by Phil Novinger and Daniel Oberlin, Florida State University.*

Let $\sum_{n=1}^\infty a_n$ be a positive term convergent series and $r_n = \sum_{m=n}^\infty a_m$, $n = 1, 2, \ldots$.

(i) Show that if $0 < p < 1$, then there is an absolute constant C_p such that

$$\sum_{n=1}^\infty \frac{a_n}{r_n^p} < C_p\left[\sum_{n=1}^\infty a_n\right]^{1-p}$$

for all such series.

(ii) Find the best possible C_p.

Solution by Nick Lord, Wolfson College, Oxford, England. The best possible constant is $C_p = \dfrac{1}{1-p}$. Since $a_n = r_n - r_{n+1}$, we have

$$\text{(1)} \qquad \frac{a_n}{r_n^p} = \left(1 - \frac{r_{n+1}}{r_n}\right)r_n^{1-p} < \frac{1}{1-p}\left[1 - \left(\frac{r_{n+1}}{r_n}\right)^{1-p}\right]r_n^{1-p} = \frac{1}{1-p}\left[r_n^{1-p} - r_{n+1}^{1-p}\right],$$

where we used Bernoulli's inequality $1 - y^\beta < \beta(1-y)$ for $\beta > 1, 0 < y < 1$. Since $r_n \to 0$ and $1 - p > 0$, so $r_n^{1-p} \to 0$ and the series $\sum(r_n^{1-p} - r_{n+1}^{1-p})$ telescopes to r_1^{1-p}. Summing (1) thus gives the result. To show that $C_p = (1-p)^{-1}$ is the best possible, let $a_n = x^{n-1}$ for $0 < x < 1$. Then

$$r_n = x^{n+1}/(1-x) \quad \text{and} \quad \sum_1^\infty a_n r_n^{-p} = (1-x)^p \sum_1^\infty (x^{1-p})^{n-1} = (1-x)^p/(1-x^{1-p}).$$

Thus

$$\left(\sum_1^\infty \frac{a_n}{r_n^p}\right)\Big/\left(\sum_1^\infty a_n\right)^{1-p} = \frac{(1-x)^p(1-x)^{1-p}}{1-x^{1-p}} = \frac{1-x}{1-x^{1-p}} \to \frac{1}{1-p}$$

as $x \to 1-$.

Also solved by K. F. Andersen (Canada), E. Badertscher (The Netherlands), H. Heining (Canada), W. Janous (Austria), R. Johnsonbaugh, O. P. Lossers (The Netherlands), V. D. Mascioni (Switzerland), W. A. Newcomb, S. Noltie, A. Tissier (France), R. Vasudera (India), J. B. Wilker (Canada), Pei Yuan Wu (Taiwan), University of South Alabama Problem Group, and the proposers.

Pei Yuan Wu notes that the problem is essentially solved in K. Knopp, *Infinite Sequences and Series*, Dover edition, p. 127, Theorem 4, where it is attributed to Dini.

Regular n and $2^k \cdot n$-gons Inscribed in a Unit Circle

E 2997 [1983, 335]. *Proposed by Irving Adler, North Bennington, Vermont.*

Let p_0 be the perimeter of an inscribed regular n-gon in a unit circle, and let d_k be the distance from the center of the circle to the side of the inscribed regular $(2^k \cdot n)$-gon. Prove that

$$\frac{p_0}{2}\prod_{k=1}^{\infty}\frac{1}{d_k} = \pi.$$

Solution I by A. A. Jagers, Technische Hogeschool Twente, Enschede, The Netherlands. Let p_k be the perimeter of the inscribed $(2^k \cdot n)$-gon, and a_k be its area. Then $a_k = \frac{1}{2}p_k \cdot d_k$, but also $a_k = \frac{1}{2}p_{k-1} \cdot 1$. Hence $p_k = p_{k-1}/d_k$ and

$$p_0 \prod_{n=1}^{k}\frac{1}{d_n} = p_k \to 2\pi,$$

the perimeter of the unit circle, as $k \to \infty$.

Solution II by J. M. Borwein, Dalhousie University, Halifax, Nova Scotia. This is a very slightly disguised form of *Viète's identity*

$$(*) \qquad \frac{\sin x}{x} = \prod_{k=1}^{\infty}\cos\left(\frac{x}{2^k}\right).$$

Indeed, it is close to his original form (see Beckmann, *A History of Pi*). We observe that

$$d_k = \cos\left(\frac{\pi}{2^k n}\right) \quad \text{and} \quad p_0 = 2n\sin\left(\frac{\pi}{n}\right).$$

The desired identity

$$\left(\frac{p_0}{2}\prod_{k=1}^{\infty}\frac{1}{d_k} = \pi\right)$$

is equivalent to $(*)$ with $x = \pi/n$.

Also solved by 48 other readers and the proposer.

ADVANCED PROBLEMS

For instructions about submitting solutions of these Advanced Problems, which should be mailed by August 31, 1986, see the inside front cover. The solver's full post-office address should be on each sheet.

6514. *Proposed by Richard Askey, University of Wisconsin, Madison.*

Show that

$$C(m,n)=\frac{(3m+3n)!(3n)!(2m)!(2n)!}{(2m+3n)!(m+2n)!(m+n)!m!n!n!}$$

is an integer for $m, n = 0, 1, \ldots$.

6515. *Proposed by Robert E. Shafer, Berkeley, CA.*

For complex μ and ν with $\mathrm{Re}(\mu)$, $\mathrm{Re}(\nu) > 0$, prove

$$\binom{\nu-1}{\mu-1}(\psi(\mu)-\psi(\nu))=\sum_{m=1}^{\infty}\frac{(-1)^m}{m}\binom{\nu-1}{\mu-1+m},$$

where $\psi(x)=\dfrac{\Gamma'(x)}{\Gamma(x)}$ and $\dbinom{\alpha}{\beta}=\dfrac{\Gamma(\alpha+1)}{\Gamma(\beta+1)\Gamma(\alpha-\beta+1)}$ (cf. this MONTHLY, 6480 [1984, 651]).

6516. *Proposed by Erwin Kronheimer, Birkbeck College, University of London, England.*

Do there exist real numbers $s_0, s_1, s_2, \ldots$, not all zero, such that each of the series

$$s_0+s_1+s_2+\cdots,$$
$$s_0+(s_0+s_1)+(s_0+s_1+s_2)+\cdots,$$
$$s_0+(s_0+(s_0+s_1))+(s_0+(s_0+s_1)+(s_0+s_1+s_2))+\cdots,$$
etc.

converges?

6517. *Proposed by Alexandru Lupaş, Facultatea de mecanică, Sibiu, Romania.*

If P_n is the sequence of Legendre polynomials, i.e.,

$$P_n(x)=\frac{1}{2^n n!}\left(\frac{d}{dx}\right)^n[(x^2-1)^n],$$

show that

$$[P_n(x)]^2-P_{n-1}(x)P_{n+1}(x)=\frac{2\cdot h_n}{\pi n(n+1)}\int_{-1}^{1}\frac{1-P_n[x^2+t(1-x^2)]}{1-t}\frac{dt}{\sqrt{1-t^2}},$$

where $h_n=\sum_{k=1}^{n}\frac{1}{k}$.

SOLUTIONS OF ADVANCED PROBLEMS

How Schlicht Is a Geometric Series?

6481 [1984, 652]. *Proposed by I. J. Schoenberg, University of Wisconsin–Madison.*

Let

$$f_n(z)=1+\frac{z}{4}+\frac{z^2}{4^2}+\cdots+\frac{z^n}{4^n}\quad(n=1,2,\ldots).$$

As a very special case of a theorem of Szegő (Theorem II in Collected Papers, vol. 2, page 171) we know that the image $f_n(U)$ of the unit circle $U=\{|z|\leqslant 1\}$ is schlicht and convex. Show that in fact

$$|f_n(z_1)-f_n(z_2)|>|z_1-z_2|/18$$

whenever $z_1, z_2 \in U$ and $z_1 \neq z_2$.

Solution by Walther Janous, Ursulinengymnasium, Innsbruck, Austria. Let z, w be in U. Since

$$|(w^k - z^k)/(w - z)| \leqslant k$$

and

$$R_n = \frac{f_n(w) - f_n(z)}{w - z} = \frac{1}{4} + \frac{w + z}{4^2} + \frac{w^2 + wz + z^2}{4^3} + \cdots$$
$$= \left(4^{n+1} - 4\frac{w^{n+1} - z^{n+1}}{w - z} + wz\frac{w^n - z^n}{w - z}\right)/4^n(w - 4)(z - 4),$$

the triangle inequality applied to the *final* expression on the right yields

$$|R_n| \geqslant \left(4^{n+1} - 4(n + 1) - n\right)/4^n \cdot 25 = \frac{4}{25} - \frac{(5n + 4)}{4^n \cdot 25} =: c_n.$$

This is better than required, and clearly best possible for $n = 2$ and n infinite.

Most solvers obtained the bound $1/18$: they applied

$$\left|\frac{1}{4} + \sum u_i\right| \geqslant \frac{1}{4} - \sum |u_i|$$

to the *penultimate* expression, made the obvious trivial estimates, and used

$$\sum_{k=2}^{\infty} k4^{-k} = \frac{7}{36}.$$

Slightly different procedures gave A. Meir and Klauss Zacharias (Berlin) the bound $1/8$; this (consider $n = 2$) is the best possible constant independent of n. Bertram Walsh pointed out that since the f_n are convex, a fact he proves directly by showing that for z in U,

$$\operatorname{Re}\left\{1 + z\frac{f_n''(z)}{f_n'(z)}\right\} \geqslant 0$$

(see, e.g., P. Duren, *Univalent Functions*, Grundlehren #259, Springer-Verlag, New York, 1983, p. 42), the best lower bound b_n is the minimum modulus of $f_n'(z)$ on U. He used this to prove that the above c_n is b_n for n even.

Johnny E. Brown's proof for $1/8$ does *not* mention n. Let D be the set of all functions g (this includes the f_n) that are analytic and univalent in the open unit disc Δ with $g(0) = 1$, $g'(0) = 1/4$, and $g(\Delta)$ convex. For $z_1, z_2 \in \Delta$ define

$$(\Phi(z_1, z_2))(g) = \frac{g(z_1) - g(z_2)}{z_1 - z_2}.$$

By "well-known results" (see, e.g., L. Brickman, T. H. MacGregor, D. Wilken, *Convex hulls of some classical families of univalent functions*, Trans. Amer. Math. Soc., 156 (1971) 91–107) the boundary of

$$V = \{\Phi(g) \colon g \text{ is in the closed convex hull of } D\}$$

is the curve (in function space)

$$\lambda(t) = \Phi(K(z, t)), \quad 0 \leqslant t \leqslant 2\pi,$$

where

$$K(z, t) = \left[1 + z\left(\frac{1}{4} - e^{it}\right)\right]/(1 - e^{it}z).$$

Hence

$$\min_{g}|\Phi(g)| = \min_{0\leqslant t\leqslant 2\pi}|4(1 - e^{it}z_1)(1 - e^{it}z_2)|^{-1} \geqslant \frac{1}{8},$$

since $z_1 = -z_2 = e^{i\phi}$ (ϕ real) are extremal for the last expression.

Also solved by K. F. Andersen, Zachary Franco, Emil Grosswald, Hong Oh Kim (Korea), Kee-wai Lau (Hong Kong), O. P. Lossers (The Netherlands), Syrous Marivani, Jean-Marie Monier (France), Victor Pambuccian (Romania), James C. Smith, Robert Vermes, Larry R. Walker, William White, an anonymous reader, and the proposer.

Nonextreme Unique Critical Points

6483 [1984, 652]. *Proposed by J. Arias de Reyna, University of Seville, Spain.*

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a differentiable function with a unique critical point at which it has a local minimum. For which values of n is the minimum necessarily an absolute minimum?

Solution by Ole Jørsboe, Technical University of Denmark, Lyngby, Denmark. The only such value is $n = 1$, for which it is clearly true. For $n > 1$ define a fifth degree polynomial

$$P = P(x_1,\dots,x_n) = x_n^2 + \sum_{i=1}^{n-1} x_i^2(1 + x_n)^3 = \sum_{i=1}^{n} x_i^2 + H,$$

where $H = H(x_1,\dots,x_n)$ consists of "higher order terms". We easily see that $0 = (0,\dots,0)$ is a local minimum for P. But P has the partial derivatives

$$\frac{\partial P}{\partial x_i} = 2x_i(1 + x_n)^3, \quad 1 \leqslant i \leqslant n - 1,$$

$$\frac{\partial P}{\partial x_n} = 2x_n + 3\sum_{i=1}^{n-1} x_i^2(1 + x_n)^2,$$

so 0 is the only critical point. Moreover, the range of P contains the range of a one-variable cubic polynomial, hence all real numbers. Thus no absolute minimum exists.

There is a similar "counterexample" in B. Calvert and M. K. Vamanamurthy, *Local and global extrema for functions of several variables*, J. Austral. Math. Soc. (A) 29 (1980), 362–368, and also some positive results. For $n = 2$ they show that polynomial counterexamples have degree at least 5, but little is known about the exact greatest lower bound for $n \geqslant 3$.

All other solutions involved some transcendental function. For $n = 2$ a solution due to David Smith is given in Philip Gillett, *Calculus and Analytic Geometry*, 2nd ed., D. C. Heath, Lexington, MA, 1984, p. 750. No fewer than three papers on this theme (together with graphics) occur in Math. Mag., 58 (1985) 146–150, authored by David Smith, J. Marshall Ash and Harlan Sexton, and Ira Rosenholt and Lowell Smylie, respectively.

Also solved by J. Marshall Ash, Robert J. Currier, Newcomb Greenleaf, M. S. Klamkin and A. Meir (jointly), Mark D. Meyerson, and James C. Smith.

An Irrationality Criterion

6484 [1985, 62]. *Proposed by A. V. Nabutovskii, Novosibirsk, U.S.S.R.*

Suppose that $x = \sum_{n=1}^{\infty} p_n/q_n < \infty$, where the p_n's and q_n's are positive integers satisfying the inequality

$$\frac{p_n}{q_n(q_n - 1)} \geqslant \frac{p_{n+1}}{q_{n+1} - 1} \quad \text{for } n = 1,2,\dots .$$

Let S be the set of indices n for which the inequality is strict. Prove that x is irrational if and only if S is infinite.

Composite of solutions by L. E. Clarke (England), John Greene, Southern Illinois University, and O. P. Lossers, Eindhoven University of Technology (The Netherlands). The given inequality is simply

$$\frac{p_n}{q_n - 1} - \frac{p_{n+1}}{q_{n+1} - 1} \geqslant \frac{p_n}{q_n}.$$

Clearly $p_n/(q_n - 1) \to 0$, so

$$\frac{p_m}{q_m - 1} \geqslant \sum_{n=m}^{\infty} \frac{p_n}{q_n},$$

with equality for large enough m if and only if S is finite. Thus S finite immediately implies that x is rational. Conversely, if $x = a/b$ is rational, set

$$r_n = \frac{a}{b} - \sum_{k=1}^{n} \frac{p_k}{q_k} = \sum_{k=n+1}^{\infty} \frac{p_k}{q_k} \leqslant \frac{p_{n+1}}{q_{n+1} - 1}$$

and

$$Q_n = bq_1 \cdots q_n.$$

Then

$$\begin{aligned} Q_{n+1} r_{n+1} &= Q_n r_n q_{n+1} - bq_1 \cdots q_n q_{n+1} \frac{p_{n+1}}{q_{n+1}} \\ &\leqslant Q_n r_n q_{n+1} - Q_n r_n (q_{n+1} - 1) = Q_n r_n. \end{aligned}$$

A decreasing sequence of nonnegative integers is eventually constant, so S must be finite and the result follows. The determination of the q_n (giver the p_n) is reminiscent of certain "greedy" algorithms that arise in the study of Egyptian fractions. (For a survey of problems related to Egyptian fractions, see P. Erdos and R. L. Graham, *Old and new problems and results in combinatorial number theory*, Monographie no. 28 de l'Enseignement Mathématique, Genève, 1980.)

Also solved by the proposer.

Approximation by Bernstein Polynomials

6485 [1985, 62]. *Proposed by Alexandru Lupaş, Sibiu, Romania.*

Given a function $f \in C[0,1]$, let $(B_n f)(x)$ denote the Bernstein polynomial

$$\sum_{k=0}^{n} \binom{n}{k} x^k (1 - x)^{n-k} f\left(\frac{k}{n}\right).$$

If $f \in C^{(2)}[0,1]$, prove that, for $0 \leqslant x \leqslant 1$, $n = 1, 2, \ldots$,

$$|(B_n f)(x) - (B_{n+1} f)(x)| \leqslant \frac{x(1-x)}{n+1} \left(\frac{1}{3n} \int_0^1 |f''(t)|^2 \, dt \right)^{1/2}.$$

Combined solution. We have the identity

$$(B_n f)(x) - (B_{n+1} f)(x) = \frac{x(1-x)}{n(n+1)} \sum_{k=1}^{n} \binom{n-1}{k-1} x^{k-1} (1-x)^{n-k} \left[f; \frac{k-1}{n}, \frac{k}{n+1}, \frac{k}{n} \right],$$

where

$$[f; x_1, x_2, x_3] = (x_3 - x_1)^{-1}\left[\frac{f(x_3) - f(x_2)}{x_3 - x_2} - \frac{f(x_2) - f(x_1)}{x_2 - x_1}\right]$$
$$= \int_0^1 H_k(t) f''(t)\, dt$$

is the second divided difference of f and

$$(x_3 - x_1) H_k(t) = \begin{cases} \dfrac{t - x_1}{x_2 - x_1} & x_1 < t \leqslant x_2, \\ \dfrac{x_3 - t}{x_3 - x_2} & x_2 \leqslant t < x_3, \\ 0 & \text{otherwise.} \end{cases}$$

Here

$$\int_0^1 H_k^2(t)\, dt = n/3,$$

so the result for the absolute value of the difference is deduced from the identity by the trivial estimate followed by the Cauchy-Schwarz inequality.

William A. Newcomb's solution uses the Euler-Lagrange equations of the calculus of variations. L. E. Clarke (England) also obtains the sharp (consider $f(x) = x^2$) estimate

$$|(B_n f)(x) - (B_{n+1} f)(x)| \leqslant \frac{Mx(1 - x)}{2n(n + 1)}$$

where

$$M = \max\{\, |f''(t)| : 0 \leqslant t \leqslant 1\}.$$

F. Schurer (The Netherlands) included in his solution a reference to O. Arama, *Some properties concerning the monotonicity of the sequence of Bernstein's interpolation polynomials and its application to the approximation of functions*, Mathematica (Cluj) 2(1960) 25–40.

Also solved by Kee-wai Lau (Hong Kong), Jean-Charles Leccia (France), Syrous Marivani, Grzegorz Rzadkowski (Poland), C. R. Selvaraj, and the proposer.

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER

TEX: A NON-REVIEW

Herbert S. Wilf
University of Pennsylvania

When I started out in mathematics, manuscripts were prepared with pencil and paper followed by typing. Lots of mathematicians still use that

method. The beauty of pencil and paper is that thoughts can flow from your head to the page without your having to deal with lots of gadgetry in between.

Pencil and paper are amazingly versatile. Insertion of new text is accomplished, perhaps, by pencilling between the lines or in the margins. Deletion can be done by the crumple-and-toss algorithm, in the case of a large deletion, or with the other end of the pencil, or by crossing things out, etc.

But there may come a time when we want to transform our handwritten document into something a little more formal; like a typewritten document, for instance.

After it has been typed, if no further changes are needed, then off it goes to its destination. But things never happen that way.

Instead we suddenly find that 13 things could have been said better and 19 other things shouldn't have been said at all. So we mark the revisions and get the manuscript retyped.

If no further changes are needed, then off it goes to its destination. But things never happen that way.

What does happen next is that we see that the whole thing would be vastly improved by a certain list of revisions, additions and deletions, but Guilt stops us from asking for those changes to be made. After all, we had our chance with the typist, then we had another chance with the typist, and surely that's all we were entitled to.

Enter the computer-based word processors.

These programs have been around for a few years now, and there are legions of novelists who don't use anything else for their manuscripts. With such programs you can keep changing and revising your text till the cows come home, and Guiltlessly too, because either you do all the work yourself, or your typist needs to type only the changes instead of retyping the whole manuscript everytime you want to rephrase something.

With a fairly inexpensive home computer and a fairly inexpensive daisy wheel printer you can run off 'rough' drafts that look good enough to photocopy and print. Word processing programs help professional writers to feel good about their works that are in progress because the rough drafts already look so good and it's so easy to make them look still better. When writers feel good about their work they write better and faster, all of which is a big plus.

That's great if in your writing you use only 26 upper and lower case letters and a handful of other characters, like writers of novels do. You get two advantages if you're one of those lucky writers. First since you ask for only a limited number of characters to print, you need only a single

standard daisy wheel on your printer, unless you want to fiddle with different type fonts, and your printouts look gorgeous. Second, for the same reason, you don't need a whole lot of special commands in your word processing program that will make a billion special characters. Your program is therefore a lot easier to use.

Therefore you get high quality printing from a cheap and simple mechanism and you get a program that's easy to use. All you have to do is to agree to want only the standard typewriter keyboard characters.

Writers of novels and short stories are delighted with their word processing systems. We mathematicians aren't delighted with ours. Our needs are not simple: we want infinitely many characters in infinitely many sizes and fonts to print, all effortlessly.

Consequently, *our* daisy wheel printers need several wheels and multiple passes to print 94 percent of our latest manuscript and then we have to handwrite in all of those extra symbols anyway. Further, if we try to avoid the above by using a dot matrix printer then we induce vertigo in our readers, to the point where many editors now reject dotty manuscripts on sight. Finally, despite the fancy word processing program that we bought, and the fancy price that it commanded, somehow everything that we want to do is just outside of its range of capability, and it doesn't *really* look good anyway, and we can't *really* run off a lot of rough drafts easily because each of them will force us to fill in by hand those 'few' special symbols and equations that the program can't handle, and if it weren't for the Guilt we'd rather go back to the secretary down the hall and the (sigh!) pencil and paper.

The problem of designing a really powerful and satisfying mathematical word processing program is formidable indeed. We mathematicians make it so by wanting that infinite character set and all those other things. Many, many programs are on the market that claim that with them you will be able to do fully satisfying mathematical manuscript preparation. The nuances of mathematical expression are so delicate, however, and the variety of symbols, spaces, fonts, etc. is so great that I have never seen one of these programs that really allows me to prepare a complex mathematical manuscript with complete confidence that it will produce gorgeous copy from virtually every mathematical thought that comes into my head.

Enter TEX.

TEX is the creation of Donald Knuth, of Stanford University, and he has described it quite thoroughly in *The TEXbook*, Addison Wesley, 1984, which is the book that this article isn't a review of.

A full review would consider a whole lot of the details of TEX, compare it with some of its competition (like Bell's *Troff*), and so forth. Instead I

want mainly to try to place the thing in some perspective, and to urge all of you to try it, because you're in for a treat.

In this non-review you'll see a rather personal statement of how this one user got hooked.

To me the most important plus for TEX is that it *feels good* to see the rough drafts. They *do* look like they've been printed already. I *won't* have to fill in any equations by hand. I *do* look forward to seeing that next draft because of how pretty I'm sure it will turn out to be. These good feelings spur me on to do whatever it is that I'm doing, faster, and I think, better, just as I said word processing systems do for professional writers. Until now we mathematicians have been deprived of that extra dose of intermediate good feelings, and now we can have them too.

TEX *can* print virtually any mathematical thought that comes into your head, and print it beautifully. If you want to let

$$x_i = \prod_{j=1}^{\max(k,i)} e^{2\pi j r s^2} \cos(ij) \quad (i = q, \ldots, n) \tag{1.3.12}$$

well then, go ahead and do it.

Your disk file won't have mysterious gaps in it because you had to hand-write things that your word processor couldn't handle into the manuscript. TEX isn't an add-on to an existing word processing program that allows it to handle a few mathematical symbols. TEX was created by a mathematician, for *us*, from the ground up, and it looks it. Using it, after having used some of those other systems, is sheer luxury.

Is it all perfect then? Well not quite. TEX is a non-user-friendly program. It belongs to the 'what-you-see-isn't-a-bit-like-what-you'll-get' school of programs. The learning process was very painful for me. For about one month, using the system several hours per week, I can recall no session in which I wasn't totally surprised by something that printed out. After three months I got moderately proficient at it, and now I can type from my head into TEX fairly fluently.

If you want to type a 'γ', for instance, you will actually type '$\gamma $', which isn't so unreasonable, but then on your screen there will appear '$\gamma $', instead of just '$\gamma$'. No doubt I'm asking for something that would cost mountains of memory, but I'd like to have it anyway.

I'm dimly aware, from scanning *The TEXbook*, that TEX is capable of doing many more things than I am personally doing with it. My present degree of proficiency is such that I can type text and equations, while having only the shakiest grip on the finer points of composing pages, alignment, boxes and glue, the innards of the macros, and so forth.

Despite my limitations and the lack of an exact, continuously visible preview on the screen (some versions allow you to get a preview of the printout, after some time delay, by asking for one), it is clear to me that TeX represents a quantum jump in our ability to convey thoughts to paper, and I strongly recommend it to all mathematicians whose hearts are stout enough to endure the learning curve.

What does TeX output look like? You're reading it now (surprise!). The Monthly has taken the output right out of my hands, photographed it, and reproduced it here. As long as your text consists only of words, the quality of the output is high but not especially better than what you can get from your home computer *cum* daisy wheel.

However, if you feel like cutting loose with an occasional

$$\begin{aligned}\left\{\sum_{\substack{i=0\\ i\text{ even}}}^{n}\binom{n}{i}2^{i^2}\right\}^r &\le \left\{2^n\sum_{i=0}^{n}2^{i^2}\right\}^r\\ &\le \left\{2^n(n+1)2^{n^2}\right\}^r\\ &= (n+1)^r2^{rn(n+1)}\end{aligned}\tag{2.7.10}$$

then TeX gives you capabilities that had heretofore been reserved to the printing trade.* You get

(a) a huge repertoire of mathematical symbols in different sizes and shapes

(b) the ability to compose your page exactly the way you would like it to appear, or if you choose to ignore this, a very carefully thought out composition scheme will be followed, with care automatically taken at page breaks, automatic hyphenation (!), and so forth

(c) lots of small touches to brighten the days of mathematicians.

When you state a theorem, for instance, you can type '\proclaim Theorem 3.6. Let f be continuous on ...' and what you'll get will be

Theorem 3.6. *Let f be continuous on ...*

What the '\proclaim' does for you is that you get 'Theorem 3.6.' boldface and flush left, and the actual statement of the theorem is set in italics.

But only a mathematician would have called that 'proclaim'.

It's time to say a little more explicitly how TeX works. As long as

* Did you notice how well nourished the big braces in (2.7.10) are?

you're typing plain English text, you just use it as you would use your own word processor. In fact you *are* using your own word processor, because TeX itself is a formatter. It takes the word processing file that you prepare using your program and converts it into a device-independent output file that will eventually cause your printer to print what you want it to print.

When you want to give a special instruction to TeX itself you type a string that begins with a backslash, like '\proclaim'. When it sees one of those ('control sequences'), TeX knows that you're talking to it and you don't want the control sequence itself printed.

There are hundreds of control sequences available to you in TeX, and you can invent hundreds more of your very own, store them and use them freely. There are control sequences that set the page length, the page width, the type font, the indentation, the type size, the presence or absence of page numbers, that handle footnotes, that leave blank spaces of various sizes, that print accents on foreign words, that print special mathematical symbols in special sizes, that ...

When you want to type a mathematical expression you first type a dollar sign, which gets you into 'math mode', then your expression, then another dollar sign and you're back in text mode. If you type 'Θ_3' then what will print will be 'Θ_3'. If you want to print 'Suppose $\sum x_i \ge 0$. Then we would have...' you would type 'Suppose $\sum x_i\ge 0$. Then we would have ...'

If you want to type a *displayed* mathematical expression, *i.e.*, one that has a whole line to itself, then you type a pair of dollar signs, your expression, and another pair of dollar signs to get back to text mode.

OK, you're interested. My comments about the learning curve haven't scared you off, and you'd like to try it out on a homework assignment for one of your classes, an exam you're about to give, a paper you're writing, or just a letter to a friend. How do you go about getting into the scene?

Certainly the first step is to buy *The TeXbook.* Read a little of it to get some of the flavor, and then try to get your hands-on experience.

To do that it would be easy for me to advise you to go to the large mainframe computer that your department owns, sit down at one of the plush terminals that are available to you, with *The TeXbook* open on your lap, type your manuscript and print it on the laser printer that's next to the water cooler in your department office.

For most mathematics departments that's just science fiction, so what do we *really* do next? If your school has a computer science department it probably has TeX running already and printing out beautiful things on its laser printer (our fiction is their reality). So cultivate friends in the computer science department. Give them a *quid pro TeX*. Bribe somebody.

Somehow get an account for yourself that will let you try TEX for yourself. Otherwise, come to an annual meeting of the Society and the Association, visit the counters where TEX is displayed, ask a lot of questions, and scheme.

Next, you're going to spend some money, either yours or your department's. TEX now can run on a maximal IBM PC/XT, or better yet, PC/AT. You could get printed output from a matrix printer, but don't, because the output wouldn't look like a printed page, and that, after all, is the name of the game. Insist on a laser printer, the prices of which have come down from astronomical to merely expensive, with more reductions in sight. For less than $9000, at the time of this writing, you or your department can be fully in business, having purchased a PC/AT, TEX itself, a laser printer, extra memory, and whatever else you need.

If the budget permits, buy one of the work stations that are now available, instead of the microcomputer. Then you'll be able to preview your printed pages just by flipping a switch, and the display will be startlingly crisp and clear. You should at least see one of these in operation before trying to convince your Dean that the continuing existence of the College is inextricably linked with your being able to install TEX in the mathematics department. I hope that your Dean gets convinced, though, because I believe that more and more we will see that TEX will become the standard against which all other systems for the preparation of mathematical manuscripts will be measured.

Then you'll be able to say

$$\prod_{j\geq 0}\Bigl(\sum_{k\geq 0} a_{jk}z^k\Bigr) = \sum_{n\geq 0} z^n \Bigl(\sum_{\substack{k_0,k_1,\ldots\geq 0\\ k_0+k_1+\cdots=n}} a_{0k_0}a_{1k_1}\cdots\Bigr)$$

or

$$\binom{n}{k} \equiv \binom{\lfloor n/p\rfloor}{\lfloor k/p\rfloor}\binom{n \bmod p}{k \bmod p} \pmod p$$

whenever you want to.

Shock Waves and Reaction-Diffusion Equations. By Joel Smoller, Grundlehren der mathematischen Wissenschaften No. 258. Springer-Verlag, New York/Heidelberg/Berlin, 1983.

BARBARA LEE KEYFITZ
Department of Mathematics, University of Houston, Houston, TX 77004

The mathematical theory of nonlinear hyperbolic conservation laws (abbreviated by Smoller as "shock waves") has long presented the casual inquirer with a formidable technical wilderness. The most basic results are stunningly difficult to establish. Even the name of the subject is somewhat mysterious. It refers to the systems of partial differential equations of a type which occurs, for

instance, when mass, energy and momentum are conserved in physical problems involving fluids in motion. A simple example is $u_t + (u^2)_x = 0$.

In the thirty-six years since Courant and Friedrichs first used the term in a mathematical sense [3], almost all aspects of the theory of a single equation have been worked out and much is now known about systems of conservation laws in a single space variable. This is a major accomplishment, even though our understanding of the latter is far from complete. For the problem of conservation law systems in several space variables even the basic existence theorems are still unknown. In this field, results are hard to obtain, and harder still to explain. In spite of the genuine importance of these equations in fluid dynamics (no, not just supersonic bombers: internal combustion engines, meteorology, elasticity and magnetohydrodynamics too) most students have simply been discouraged by the inaccessibility of the subject.

On the other hand, if you put "reaction-diffusion equations" in the title of your colloquium, you are guaranteed a large audience. Perhaps young mathematicians are attracted by the ecological applications (who can fail to relate to rabbits diffusing rapidly to escape predatory foxes) as much as by the unified and elegant analytical background. Whatever the reason, they come to listen and they stay to contribute: Fife's monograph of 1979 [5] lists over two hundred research articles, most of them recent.

Is it not ironic, then, to find these two fields combined in a single textbook? I'm not sure that, beyond the fundamental unity of all mathematics, the study of one contributes to an understanding of the other. It must be said, however, and I think this justifies the whole book, that Smoller manages to bring the same level of excitment and enthusiasm to his treatment of conservation laws as he does to reaction-diffusion equations. And at the elementary level, which is the main emphasis of the book, the two subjects have much in common. They are treated separately in the book, but an introductory section shows how both are related to the fundamentals of partial differential equations. Smoller's text is organized like a sandwich: first an overview of the theory of linear hyperbolic and parabolic equations; then the meat (two kinds): reaction-diffusion equations, and the theory of shock waves. The last quarter of the book wraps the two together by another unifying theme: the Conley Index, which has been applied to many problems in both fields. In fact, although all four topics have been more than competently dealt with in separate monographs [2], [5], [6], [7], [8], this book is unique in combining them. It is also unique in being the only complete and self-contained reference on the elementary aspects of the two main subjects. As such it has already become a standard reference work.

The decision to present the common mathematical ancestry of the subjects incurs a cost: many of the physical models that have motivated and guided the study of each field are necessarily slighted. The book includes a chapter on topology (both differential and algebraic), but none on fluid dynamics. The field of combustion, important for both subjects, is nowhere mentioned. Nor is there a discussion of computational methods, which is a pity since the interaction between numerical approximation and theory has been particularly fruitful in the field of conservation laws.

On the other hand, the mathematical framework is ideal for explaining to the novice the origin of the technical difficulties of conservation laws. The subject, one might say, is still searching for its roots. In the classical theory of partial differential equations, equations are divided into types (hyperbolic, parabolic, or elliptic) on the basis of algebraic criteria. For example, an equation is hyperbolic if matrices formed from the coefficients have real eigenvalues. The algebraic conditions turn out to have profound analytic consequences. For example, when one of the independent variables represents time, then only in the hyperbolic case does the initial value problem have a solution which depends continuously on the data, even for a short time. For simple models of physical problems, this is consistent with physical experience: equations like the wave equation ($u_{tt} - c^2 u_{xx} = 0$) which model time dependent physical processes are indeed hyperbolic and the mathematical solutions look like the physical ones, even when the time variable becomes very large.

But make the problem ever so slightly more complicated—let the sound speed, c, in the wave equation depend on the pressure, u, for instance—and the classical mathematical theory predicts that solutions will exist only for a short time, after which they spontaneously develop discontinuities. What do we do? Surely we have a responsibility, after invoking physics to justify one mathematical theory, to respond to this appeal to make mathematical sense out of another physically reasonable problem. So, for thirty years, mathematicians have hammered out the foundations of the subject with a succession of mathematical and physical tools. First they introduced "weak solutions": discontinuous functions that satisfy the equation in an averaged sense. But this admitted too many solutions: uniqueness was lost. They then introduced frankly physical criteria to distinguish admissible weak solutions. Then mathematicians had to run through the cycle one more time to make these new notions rigorous. Right now, the most exciting development in conservation laws is the use of a deep tool of nonlinear functional analysis, known as the method of compensated compactness, which shows extraordinary promise of revealing the correct mathematical framework. Throughout, the interplay of physical ideas with hard analysis has been at the heart of the subject: the technical nature of research in conservation laws is built into the questions that mathematicians want answered.

The mathematical motivation for research in reaction diffusion equations is rather different. (The equation $u_t = bu_{xx} + f(u)$ is an example of a reaction-diffusion equation.) Models from mathematical genetics and from chemical engineering and combustion presented challenging examples which were studied even before the term "reaction-diffusion equation" was coined. Perhaps because the fundamental questions, while still mathematical, are considerably less technical than in shock wave theory, there has been substantially more cooperation between mathematicians and engineers. For example, Aris's treatise [**1**], while written for chemical engineers, is a first-rate source of mathematical problems: multiplicity and stability of steady states, and stable types of transient behavior such as travelling waves. Much current research in the broad area now known as reaction-diffusion equations centers about such problems, especially the last one. About a decade ago, Kopell and Howard suggested a new way of looking at qualitative properties of solutions of the complicated equations modelling chemical and biological phenomena. Rather than looking at particular, complex systems, they proposed to construct classes of equations that might have the desired properties. Thus was born the subject of reaction-diffusion equations. Another recent development is the study of so-called dissipative structures which evolve as a consequence of differential diffusion rates in systems of reaction-diffusion equations. This area, discovered independently by biologists, astrophysicists and combustion scientists, is now ripe for mathematical study. It will necessarily involve the interaction of reaction and diffusion in many subtle ways.

Recent interest in mathematical ecology has suggested somewhat different directions of research, and it is this area which interests Smoller the most. After a very effective presentation of two of the basic tools in reaction-diffusion equations, comparison theorems and linearized stability, his treatment concentrates at some length on the construction of invariant regions, an important concern in ecology models. The reader who masters the section on the Conley Index will certainly learn something about constructing travelling waves, although the book does not emphasize their importance in mathematical genetics and combustion problems. The chapter on the study of stability via linearization represents, as far as I know, the only attempt at a readable, self-contained account aimed at applications in reaction-diffusion equations. For this the author merits the gratitude of everyone trying to learn or teach this subject. And a chapter on bifurcation theory, with an example, introduces the student to the way bifurcation theory is used in more advanced applications. However, the applications themselves are missing. Perhaps there was just no more room in the book, even for some of the "textbook" examples. For those the reader should consult Fife [**5**], whose treatment is in a sense complementary to Smoller's: here one will find several suggestive applications and an extremely generous annotated bibliography which although now a few years old, surpasses Smoller's for completeness and for indicating directior

for further reading.

The same point, that limitations of space have prevented Smoller from recording recent results, applies also to the subject of conservation laws. The student who is captivated by Smoller's extremely clear and readable presentation may be surprised to discover how much more there is to learn than what is here. There is a complete discussion of the theory of a single conservation law, and a good introduction to systems, including what is still the most important result of the entire subject: a fundamental theorem, proved by Glimm in 1965, which answers the long-time existence question. Glimm's proof is valid only under the hypothesis that the initial data is almost constant, and extending this result is another consuming passion of current research. Smoller's coverage of the development of the subject since Glimm's theorem is sketchy, however. For reference to topics that should, perhaps, have been mentioned in the text, the reader should consult the papers in [**9**].

As I have tried to make clear, the strength of Smoller's book is in its introductory material. While contemplating the book, I grumbled to a colleague that I thought books in the Grundlehren series were supposed to be comprehensive. "Oh no", he replied, "everybody knows they are just supposed to be incomprehensible." It is delightful to state that Smoller's book fails on this count.

References

1. R. Aris, The Mathematical Theory of Diffusion and Reaction in Permeable Catalysts, vols. 1, 2, Clarendon Press, Oxford, 1975.

2. C. C. Conley, Isolated Invariant Sets and the Morse Index, CBMS Regional Conf. Series in Math., 38, Amer. Math. Soc., Providence, 1978.

3. R. Courant and K. O. Friedrichs, Supersonic Flow and Shock Waves, Wiley, New York, 1948.

4. B. Dacorogna, Weak Continuity and Weak Lower Semi-Continuity of Non-Linear Functionals, Lecture Notes in Mathematics No. 922, Springer-Verlag, 1982.

5. P. C. Fife, Mathematical Aspects of Reacting and Diffusing Systems, Lecture Notes in Biomathematics No. 28, Springer-Verlag, New York, 1979.

6. G. B. Folland, Introduction of Partial Differential Equations, Princeton Univ. Press, 1976.

7. P. D. Lax, Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves, Conf. Board Math. Sci., 11, SIAM, 1973.

8. C. S. Morawetz, Nonlinear Waves and Shocks, Tata Inst. Notes, Springer-Verlag, New York, 1981.

9. J. A. Smoller, ed., Nonlinear Partial Differential Equations, Contemp. Math., 17, Amer. Math. Soc., Providence, 1983.

Geometry: *A High School Course*. By Serge Lang and Gene Murrow. Springer-Verlag, New York, 1983. xxiii + 470 pp. $24.00.

Seymour Schuster
Department of Mathematics, Carleton College, Northfield, MN 55057

The teaching of geometry, particularly at the secondary level, has been troubling mathematicians for at least three decades. This topic constituted one of the main themes of the section of the 1958 Edinburgh Congress that was administered by the International Commission on Mathematical Instruction and since then there have been numerous conferences, equally numerous reports, and even more numerous talks and papers dealing with all aspects of the problem. Among the many contributors to the dialogue have been distinguished mathematicians such as H. Behnke, J. Dieudonné, J. Freudenthal, E. Moise, and R. Thom. Yet, it appears that little progress has been made, at least in the U.S.; in fact, there are many who contend that we are worse off now than we were three decades ago in the sense that students entering the calculus sequence today know less useful geometry than did their counterparts of one or two generations ago.

The question of why geometry education suffers so is certainly complex, for an answer entails sociological and political considerations as well as the more obvious mathematical and educational ones. In this short review article, it is only possible (and appropriate) to address a few of the latter, and this must be done briefly. We therefore confine ourselves to three assertions about the troubles with school (and perhaps collegiate) geometry teaching: (1) there is a misplaced emphasis on formal structure and rigor; (2) there is near disregard of geometry in the physical universe and the world of art; and (3) there exists, among many teachers and authors, an over-zealous devotion to a single geometric method.

Misplaced emphasis on formal structure. Historically, Euclidean geometry (taught at the secondary level in the U.S.) has been the vehicle for teaching axiomatic reasoning. This mistake has caused—and continues to cause—heavy damage to geometry education.

To begin with, it is a poor pedagogical idea. As far back as 1765, it was recognized by A. -C. Clairaut [2] that: "Although geometry is by itself abstract, we have to acknowledge that the difficulties met by the beginners can rather often be ascribed to the way in which its ordinary elements are taught. One always starts with a great number of definitions, postulates, axioms and preliminary principles that seem to promise nothing but dry facts to the reader. The propositions that follow then do not fix the mind on more interesting objects, and as these are also difficult to conceive it commonly happens that the beginners get bored and fed up before having gotten any distinct idea of what one is going to teach them." In more recent times, Clairaut's banner has been picked up by his countryman, René Thom, who in [4] and [5] has made profound analyses of the modernization of school curricula. In the latter, he states that "the real problem of mathematics teaching is not that of rigor, but the problem of the development of 'meaning,' the 'existence' of mathematical objects," and adds that the true lesson to be extracted from Hilbert's axiomatics is that "one accedes to absolute rigor only by eliminating meaning... ."

Burdening the course in elementary geometry with the chore of teaching formal structure, deductive reasoning, and rigor is not only unfair to students and teachers, and not only does it have the unfortunate effect of compromising a beautiful subject, but it is a responsibility placed on the course that has not been and cannot be fulfilled. The recognition that it has not been fulfilled is precisely one of the reasons for the many conferences on geometry teaching. It cannot be fulfilled because the subject is far too complicated to axiomatize honestly, and if an honest axiomatization were to be presented, then the course would surely be killed with theorems that have the excitement of "If M is a midpoint of segment AB, then M is between A and B," and "A segment cannot have two distinct midpoints." One need only glance at H. G. Forder's "The Foundations of Euclidean Geometry," or at [1] and [3], to see the difficulties inherent in presenting a rigorous treatment of Euclidean geometry at any level, much less in the secondary schools.

Disregard for geometry in the physical universe. While intriguing problems of elementary geometry often capture the imagination of some students, there are many more who would be captured by seeing the subject as the intellectual outgrowth of scientific and technological concerns. The abstract character of geometry should be seen as a means of simplifying a highly complex world that is, most of the time, too difficult to deal with if one were to consider all aspects of the phenomenon being studied. We idealize that world, and we make discoveries by analyzing, experimenting, and making deductions; and we reach deeper insights when we generalize. The result is a language and tool for comprehending and communicating about the geometric aspects of the physical world.

Unfortunately, the rich worlds of art, architecture, biology, chemistry, geology, and physics are often overlooked as sources of interesting geometry and motivation for developing the subject in the very first place. A particular tragedy that has been, in part, a consequence of the alienation of geometry from all worlds other than philosophy is that training in three-dimensional thinking is nearly absent from the school curriculum in the U.S. This state of affairs is decried not only by

teachers of multivariate calculus, but even more vociferously by chemistry, geology and physics professors.

Overzealous devotion to a single geometric method. Resistance by traditionalists to the use, in school geometry, of any method other than the synthetic method of Euclid is probably what led to Dieudonné's cry: "A bas Euclid." On the other hand, some of the so-called "reformers" of the 1950's and 1960's were just as rigid in their attempts to convert school geometry into analytic geometry. In more recent years, we've seen a new crop of reformers who wish to make Euclidean geometry part of linear algebra.

All of these rigid schools are wrong mathematically, as well as pedagogically, speaking. Even without the advantage of a *genus-differentia* definition of geometry, we all recognize geometric problems when we see them (except in some very esoteric cases). We attack these problems with a variety of techniques: synthetic, algebraic, analytic (including calculus), group-theoretic, etc. Sometimes one technique is impossible for that particular problem. For example, proving that the angle bisectors of a triangle meet in a point is trivial by synthetic methods and vicious by analytic methods; analytic techniques are natural for studying the focal properties of conic sections while synthetic methods are not; linear algebraic (vector) methods are ideal for a host of affine problems such as proving that the medians of a triangle are concurrent, but just try to use linear algebra to prove the Sylvester-Gallai Theorem: *Given n noncollinear points in the plane, then there exists a line through exactly two.* The abundance of such examples makes it plain that textbooks and teachers do an enormous disservice to students when they extol the virtues of one method and "badmouth" others.

Of course, not all teachers and textbooks are guilty of the three aforementioned misdemeanors. Our comments are offered only in a broad statistical sense about prevalent philosophy, influence and practice.

Now, how does the book under review stand up to the criteria implicit in the foregoing discussion? The answer is that it is excellent with respect to (1) and (3), but only fair on (2).

Lang and Murrow have written an unpretentious textbook. There is not the slightest hint that their book is to serve as a model of logic or as a school version of Hilbert's, Veblen's, or Enriques' work. Yet, there are wholesome discussions of assumptions for logical argument and a gentle introduction to logical proof. Almost always, the assumptions (i.e., postulates) are agreed upon only after the students have been asked to experiment and discover their plausibility. While the authors impress upon the reader the need for precision of language, they refrain from being finicky and pedantic, and go so far as to describe accurately and exploit explicitly the "abuse of language" that is the *modus operandi* of working mathematicians. On some occasions, however, they become a bit too casual, falling into inconsistencies. For example, at one point in their description of a ray, the ray does not contain its origin but at another it does; they define a segment to be open with the embarrassing consequence that a triangle does not contain its vertices; their definition of prism is that of a right prism, so a parallelepiped is always rectangular, contrary to their figures and discussion. (These are minor slips that are not unusual in first editions.) On other occasions, they skillfully avoid the pedagogically damaging mess that the devotees of rigor create. An example of such is their treatment of angle, in which they boldly use the full power of separation without fuss, giving the student all that is necessary for the interesting geometry that follows. For beginners (and perhaps for others, as well), learning interesting geometry is much preferred to the tiresome details of separation.

One of the admirable accomplishments of the authors is their use of a variety of geometric methods. They have chosen their topics (not always to the taste of the reviewer) and they have treated these in the most natural way, or else they have chosen the geometric methods and they have used these on the most appropriate topics. In addition to synthetic methods, students will learn the rudiments of coordinate methods, vector methods (through dot products), and the methods of transformations. Their use of transformations is particularly skillful. While most textbooks focus primarily on isometries, making some use of reflections, and some go further in

treating similarities, Murrow and Lang begin with dilatations and make heavy use of these. They analyze the effect of dilatations—particularly shears—on area and length, and then use their results together with intuitive limit arguments in the treatment of length, area, and volume. Hence, the students are also introduced to some notions of calculus as these might have been understood by Eudoxus and Archimedes.

In the attempt to relate the study of (abstract) geometry to the (concrete) geometry of the surrounding world, the authors have offered little that is interesting or exciting. They do provide a modest number of applications, occasionally as illustrations in the text, but most often in exercises. Perhaps it would have been to their advantage to begin the book with some of these applications, thus motivating the abstractions of their development.

One final weakness to be mentioned is the dearth of non-routine exercises. A good number of the exercises are good in that they have been carefully planned to foreshadow ideas that will be studied later. However, one is struck by the fact that this more-than-respectable introductory textbook does not exploit the well-known power of its subject to seduce bright students with challenging problems.

The availability of mathematically and pedagogically good textbooks is a necessary, but certainly not sufficient, condition for the improvement of geometry education. Lang and Murrow have made a substantial contribution toward satisfying that necessary condition. One gets the impression that they might have wanted to write a more revolutionary book. We should be thankful that they did not, for revolutionary works are neither understood nor given sympathetic consideration by boards of education.

References

1. R. Brossard, Metric postulates for space geometry, this MONTHLY, 74 (1967) 777–788.

2. A. -C. Clairaut, Eléments de Géométrie, Gauthier-Villars, Paris, 1920.

3. S. Mac Lane, Metric postulates for plane geometry, this MONTHLY, 66 (1959) 543–555.

4. R. Thom, "Modern" mathematics: an educational and philosophical error?, American Scientist, 59 (1971) 695–699.

5. ______, Modern Mathematics: Does it Exist?, Developments in Mathematics Education, Proceedings of the Second International Congress on Mathematics Education, Cambridge University Press, 1973.

Wheels, Life and Other Mathematical Amusements. By Martin Gardner. W. H Freeman and Company, New York, 1983. ix + 260 pp. $15.95.

DAVID A. KLARNER
Department of Computer Science, University of Nebraska, Lincoln, NE 68588

This is Martin Gardner's tenth collection of Mathematical Games columns that he wrote for *Scientific American*. As usual with these collections, they are corrected and brought up to date by reports of the discoveries made by his enthusiastic readers. Regrettably to many, Gardner has stopped writing his column regularly, and it is missed. After his 65th birthday on 21 October 1979, at least two mathematical works have been dedicated to him: *The Mathematical Gardner*, a collection of papers written by thirty-three mathematicians, and *Winning Ways*, a two-volume work on two-person games. The dedication in the latter says very eloquently, "To Martin Gardner, who has brought mathematics to more millions than anyone else."

How do we account for Gardner's wonderful success? What can we teachers and writers of mathematics learn from him? In this brief appreciation of his latest book, I shall attempt to point out the good features of Gardner's writing and the reasons for his success. One realizes, as one gets to know him through his writing, that he is humble, generous, culturally broad, and very honest. Many of these good qualities are exemplified when he is fighting mad about some hoax or

fad—armed with the mighty pen!

Here are the best features of Gardner's writing: His taste and choice of topics; the simplicity of the exposition, and, in particular, the use of examples versus generality; the generous citing of other people's ideas; enlivening cross-reference to art, literature, philosophy, politics; the use of humor, and other things that are fun.

When we are asked to give a mathematical talk to a general audience, many of us feel we have little choice of the topic. As a known expert on hemi-demi flipoids, that's the subject, and one has to cut through all the background and motivation to get right to the most important part of the subject, namely, what one has just found out a couple of months ago! Actually, it is better to tell the audience what was attractive to us in the first place. Gardner homes-in on these attractive ideas. There are several examples in his latest book; for instance, it is no exaggeration to say that a substantial motivation for modern algebraic number theory lies in the Diophantine equation mentioned in Fermat's Last Theorem. (This is the subject of Chapter Two.) The higher cardinals of set theory are discussed in the fourth chapter. Quite often, Gardner begins with an intriguing puzzle, problem, or game and leads us to an interesting deeper theory. For example, Conway's game of Life (the subject of the last three chapters) leads into the modern theory of computation. It turns out that one can use Life configurations to imitate the computation done by a Turing machine! This is characteristic of Gardner's exposition. One begins in a playful mood thinking about some puzzle or game, and the analysis leads to a deeper idea. But none of the fun is lost in this intellectual pursuit.

The exposition is as simple as possible. Gardner avoids using abstract symbols and notation. One seldom sees complicated notation or terms invented to express one idea just once or twice. Sometimes one sees in other mathematical writings a term or notation that has been introduced just to simplify one subsequent sentence, then it is never used again. Gardner's writing has almost no formulas, and a great deal is expressed in plain English familiar to all. One technique for avoiding notation is to explain the general case by means of a well-chosen example. Most of us discover and come to understand ideas that are later expressed in very general terms by means of instances that are very concrete to us. Only later do we notice that the good idea which brought about the solution is an idea which applies to a more general situation. Unfortunately, it is the abstract and general form that is recorded in the end. (This tendency is probably caused by a problem of ego. One wants to avoid the one-upmanship game played by many thinkers. Proving a more general statement in a broader context is thought to be one-up from the original statement. Generalization is often something that can be done automatically compared to the rather difficult creative act involved in the initial discovery.) From a logical point of view, generality is the desired goal, but from a learning point of view, it is not the way to begin. One should begin with the concrete and move toward the general. Better yet, if you have a generous spirit, as does Martin Gardner, you use a cunningly chosen example to allow your reader to discover the general case without further help. You let the reader feel he has gone one-up beyond what has been written. What an excellent feeling to produce in a reader! It does not matter that the reader's observation is known to the writer or other experts; for the moment, through his own discovery, an idea has become the personal property of the reader. Allowing, even encouraging, this feeling is not fashionable. It seems dishonest to trick the reader, and besides the writer may seem a bit foolish not to have said the deepest consequence of his ideas. Indeed, it takes a generous spirit to leave unsaid the most general truth one has seen. Sometimes it is better to share this satisfaction with one's reader.

Having touched on the subject of generosity, let us pursue it seriously for a moment. An expositor or teacher goes about the job knowing that a gift is being given. One can give generously or selfishly. Martin Gardner gives in a most generous way, and he does it without involving his own ego. He selects the best and most appealing aspects of his subject, spending a lot of time sorting, looking for the nicest things to give away. Furthermore, he is lavish in citing the work of others. How nice it is to see that mathematics is truly a joint effort, and not just the invention of a single author!

Of course, Gardner links mathematics to all sorts of human endeavor—art, literature, science are all mixed in. Mathematics does not have to be separate from the pleasant things around us. Mathematics is part of these things, anyway, and Gardner keeps us aware of this. I have been reading Gardner since high school and admire him more as time goes on. His latest book should be a pleasure to every reader—beginner and old salt, abstract and applied. There is something here for everyone.

LETTERS TO THE EDITOR

For instructions about submitting letters for publication in this department see the inside front cover.

Editor:

In the paper "Trees and power-sums" (this MONTHLY, 92 (1985) 328–331), A. H. Stone proves the identity

$$\frac{1}{2}\sum_{r=1}^{n-1}\binom{n-2}{r-1}r^{r-2}(n-r)^{n-r-2}=n^{n-3} \tag{1}$$

by counting trees in two different ways, and remarks that it would be interesting to have a direct elementary proof of the identity.

Let

$$A_m(x,y;p,q)=\sum_{k=0}^{m}\binom{m}{k}(x+k)^{k+p}(y+m-k)^{m-k+q}.$$

Identities for these sums are called *Abel identities* after Abel's generalization of the binomial theorem [**1**]

$$A_m(x,y;-1,0)=x^{-1}(x+y+m)^m. \tag{2}$$

The identity (1) is the case $m=n-2$, $x=y=1$ of the Abel identity

$$A_m(x,y;-1,-1)=(x^{-1}+y^{-1})(x+y+m)^{m-1}. \tag{3}$$

We can derive (3) from (2) with the recurrence

$$A_m(x,y;p,q)=A_{m-1}(x,y+1;p,q+1)+A_{m-1}(x+1,y;p+1,q),$$

which follows from the recurrence $\binom{m}{k}=\binom{m-1}{k}+\binom{m-1}{k-1}$, together with the symmetry $A_m(x,y;p,q)=A_m(y;x;q,p)$.

Abel's formula (2) can be proved directly, using the fact that if $p(k)$ is a polynomial in k of degree less than r, then the rth difference

$$\sum_{k=0}^{r}(-1)^{r-k}\binom{r}{k}p(k)$$

is zero. We have

$$\begin{aligned}A_m(x,y;-1,0)&=\sum_{k=0}^{m}\binom{m}{k}(x+k)^{k-1}\sum_{j=0}^{m-k}\binom{m-k}{j}(x+y+m)^j(-x-k)^{m-k-j}\\&=\sum_{j=0}^{m}\binom{m}{j}(x+y+m)^j\sum_{k=0}^{m-j}(-1)^{m-j-k}\binom{m-j}{k}(x+k)^{m-j-1}\end{aligned}$$

$$= x^{-1}(x+y+m)^m$$

since $(x+k)^{m-j-1}$ is a polynomial in k of degree $m-j-1$ except when $j=m$.

Riordan [3, pp. 18–23] gives a good account of Abel identities. Proofs of (2) and (3) can also be found in Lovász [2; p. 19, problem 44(b); solution, p. 173].

References

1. N. H. Abel, Beweis eines Ausdruckes von welchem die Binomialformel ein einzelner Fall ist, J. Reine Angew. Math., 1 (1826) 159–160.

2. L. Lovász, Combinatorial Problems and Exercises, North-Holland, 1979.

3. J. Riordan, Combinatorial Identities, Wiley, 1968.

Ira Gessel
Department of Mathematics
Brandeis University
Waltham, MA 02254

Editor:

I have read with interest the article "Tricks or treats with the Hilbert matrix" by Man-Duen Choi in the May 1983 issue of the MONTHLY. Allow me to add a trick of my own.

The Hilbert matrices are important in numerical analysis. They are notoriously "ill-conditioned". This has the effect that in a system $AX = B$, with A a Hilbert matrix, small changes in B may produce huge changes in X. (Thus such systems provide severe tests of programs designed to invert matrices or solve linear systems.) The following example is particularly dramatic.

Let A_3 be the 3×3 Hilbert matrix and consider the system $A_3 X = B_0$, where

$$B_0 = {}^t(11/6, 13/12, 47/60) = {}^t(1.8333\ldots, 1.08333\ldots, .78333\ldots).$$

This has the exact solution $X = {}^t(1,1,1)$. Now consider the system $A_3 X = B_1$, where $B_1 = {}^t(1.8, 1.1, .78)$. This has the exact solution $X = {}^t(0, 6, -3.6)$!

Steven H. Weintraub
Department of Mathematics
Louisiana State University
Baton Rouge, LA 70803-4918

Editor:

Your readers may be interested in knowing that Francisco Faà di Bruno (probably best known to them for the formula for the nth derivative of a composite function; note that "Faà di Bruno" was the family name) has taken the first step toward sainthood (not, alas, in recognition of his mathematics). Having attained the rank of Venerable, he may now (if I understand correctly) be prayed to for guidance in combinatorics. I don't know of any other mathematician who has got so far.

Readers who are at ease with Ecclesiastical Latin may check the facts in Acta Apostolicae Sedis 63 (1971) 788-790.

R. P. Boas
Department of Mathematics
Northwestern University
Evanston, IL 60201

JOIN MAA TODAY!

Send this application to: MAA Membership Department
1529 18th Street, N.W.
Washington, D.C. 20036

Name ______________________________

Mailing Address ______________________________

______________________________ Zip ______________

Employer/School ______________________________

Position (Rank) ______________________________

Employer's City/State ______________________________

Highest Earned Degree ______________ Year Degree Earned ______________

Institution Awarding Degree ______________________________

Month/Year of Birth ______________________________

Have you been a member of the MAA before? ☐ **Yes** ☐ **No**

Payment Enclosed $______________ *U.S. Funds Only* (see below)

Your membership may begin on July 1 or January 1. You may elect to receive any of the listed combinations of the three MAA journals:

THE AMERICAN MATHEMATICAL MONTHLY **(M)**
MATHEMATICS MAGAZINE **(G)**
THE COLLEGE MATHEMATICS JOURNAL **(J)**

Please circle the appropriate box corresponding to your initial membership period and the selection of journals you wish to receive. (Rates guaranteed for the indicated periods only.) Subscription prices are included with dues (see **).

	M	**G**	**J**	**M + G**	**M + J**	**G + J**	**M + G + J**
Student Membership							
1 year (Jan.-Dec. 1986)	$24	$18	$20	$30	$32	$26	$38
1½ year (July 1986-Dec. 1987)	$36	$27	$30	$45	$48	$39	$57
Regular Membership							
1 year (Jan.-Dec. 1986)	$40	$28	$31	$52	$55	$43	$67
1½ year (July 1986-Dec. 1987)	$60	$42	$46.50	$78	$82.50	$64.50	$100.50

*Student Membership available to high school and undergraduate students and to students regularly enrolled in graduate study at least half time. Enclose letter from school or department official confirming status. Student rates apply to unemployed persons under certain circumstances.

**Annual dues include subscription prices as follows: Reg. Memb. $24 (M), $15 (J), $12 (G), $1 (Focus). Std. Memb. $12 (M), $8 (J), $6 (G), $1 (Focus).

METHOD OF PAYMENT ☐ Check enclosed ☐ VISA ☐ MasterCard ☐ Bill me

Card Number |_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_| Expiration Date |_|_|_|_| *(month/year)*

Interbank Number |_|_|_|_| *(MasterCard only—located above name on card)*

SIGNATURE ______________________________

Carus Mathematical Monograph #21

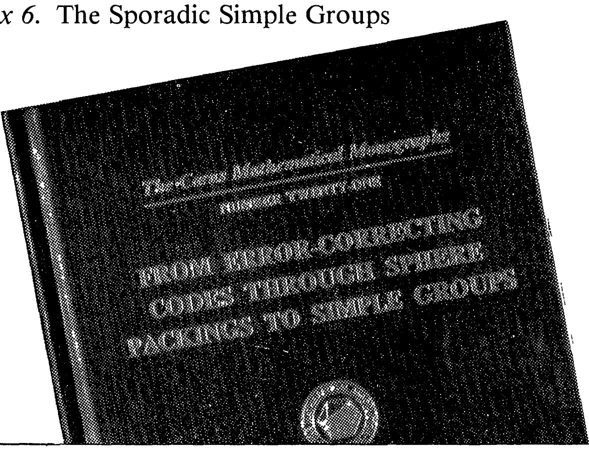
From Error Correcting Codes through Sphere Packings to Simple Groups,

by Thomas M. Thompson
224 pp. Hardbound
List: $24.00 MAA Member: $18.50

Two of the most fascinating problems to challenge mathematicians in recent years concern the construction of data transmission codes that can correct errors introduced by static and the search for efficient ways to pack ping-pong balls into a box. Can one design the best error-correcting codes? Can one find the most efficient sphere packing?

Therein lies a fascinating story which is told with great skill and clarity in this important addition to the Carus Mathematical Monograph series. The author has packed (sic) into 175 pages all of the basic mathematical ideas of this saga woven into a gripping historical account of the journey from error-correcting codes to sphere packings to simple groups.

Table of Contents



From The Mathematical Association of America

VOLUME 6—DOLCIANI MATHEMATICAL EXPOSITIONS

MAXIMA AND MINIMA WITHOUT CALCULUS,

by Ivan Niven
xv + 323 pp. Hardbound
List: $28.50 MAA Member: $21.30

Ivan Niven, distinguished author of several books on number theory and probability, has compiled the basic elementary techniques for solving maxima and minima problems. Since many books and courses already cover calculus and linear programming techniques, the author deliberately omits these areas from his discussions and concentrates instead on methods in algebra and geometry not so widely known. These methods are organized according to the mathematical ideas used, and many chapters can be read independently without reference to what precedes or follows. Some of the problems presented are left for the reader to solve with sketches of solutions given in the later pages.

The book is written for an audience at or near the maturity level of second- or third-year college students with a good working knowledge of precalculus mathematics. Although calculus is not a prerequisite, a prior knowledge of that subject will enhance the reader's comprehension. An excellent sourcebook in the area of maxima and minima, the book will serve as a textbook or as enrichment material for the talented undergraduate.

The main topics covered are:

- **Simple Algebraic Results**
- **Elementary Geometric Questions**
- **Isoperimetric Results**
- **Basic Trigonometric Inequalities**
- **Polygons Inscribed and Circumscribed**
- **Ellipses**
- **The Bees and Their Hexagons**
- **Further Geometric Results**
- **Applied and Miscellaneous Problems**
- **Euclidean Three-Space**
- **Isoperimetric Results not Assuming Existence**
- **Postscript on Calculus**

THE MATHEMATICAL ASSOCIATION OF AMERICA

Order from:
The Mathematical Association of America
1529 Eighteenth St. NW
Washington, DC 20036

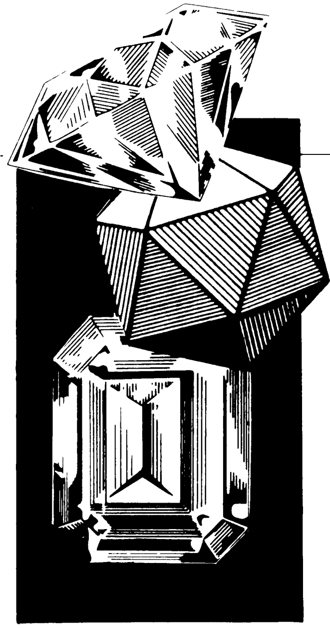
The Mathematical Association of America

CARUS MATHEMATICAL MONOGRAPHS

Students and teachers of mathematics as well as nonspecialists and workers in other fields, will all appreciate the expository excellence of these monographs. A wide range of topics in pure and applied mathematics are covered, each one written by a renowned expositor in the field. A glance at the list of titles of the books and the list of authors will convince you of the quality of this respected series of monographs. Eighteen titles are currently available:

		List Price	MAA Member
No. 1.	**Calculus of Variations**, by G. A. Bliss	$19.00	$14.00
No. 2.	**Analytic Functions of a Complex Variable,** by D. R. Curtiss	19.00	14.00
No. 4.	**Projective Geometry**, by J. W. Young	19.00	14.00
No. 6.	**Fourier Series and Orthogonal Polynomials,** by Dunham Jackson	19.00	14.00
No. 8.	**Rings and Ideals,** by N. H. McCoy	19.00	14.00
No. 9.	**The Theory of Algebraic Numbers** (Second edition) by Harry Pollard, and H. G. Diamond	19.00	14.00
No. 11.	**Irrational Numbers,** by Ivan Niven	19.00	14.00
No. 12.	**Statistical Independence in Probability, Analysis** and Number Theory, by Mark Kac	19.00	14.00
No. 13.	**A Primer of Real Functions** (Third edition), by R. P. Boas, Jr.	19.00	14.00
No. 14.	**Combinatorial Mathematics,** by H. J. Ryser	19.00	14.00
No. 15.	**Noncommutative Rings,** by I. N. Herstein	19.00	14.00
No. 16.	**Dedekind Sums,** Hans Rademacher and Emil Grosswald	19.00	14.00
No. 17.	**The Schwarz Function and its Applications,** by P. J. Davis	19.00	14.00
No. 18.	**Celestial Mechanics,** by Harry Pollard	19.00	14.00
No. 19.	**Field Theory and its Classical Problems,** by Charles Hadlock	24.00	18.50
No. 20.	**The Generalized Rieman Integral,** by R. M. McLeod	21.00	15.20
No. 21.	**From Error-Correcting Codes through Sphere Packings to Simple Groups,** by Thomas M. Thompson.	24.00	18.50
No. 22	**Random Walks and Electric Networks,** by J. Laurie Snell and Peter Doyle.	24.00	18.50

ORDER FROM:
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036



Mathematical Gems III

by Ross Honsberger

260 pp. Hardbound, 1985.

ISBN-0-88385-313-2

MAA Member: $21.00

List: $27.00

This book is packed with gems from elementary combinatorics, geometry, number theory, and probability. It will bring the enjoyment of good mathematics to the student as well as the teacher.

Glance at the Table of Contents and order now!

Order from:

The Mathematical Association of America

1529 Eighteenth Street, N.W.

Washington, D.C. 20036

1529 Eighteenth Street, N.W.
Washington, DC 20036

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 93, Number 5 May 1986

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. (This is necessary for indexing purposes.) Send all main articles, letters to the editor, and editorial correspondence to the editor-elect: H.S. WILF, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Contributions to the other four special departments (Notes, Teaching of Mathematics, Unsolved Problems, and Problems and Solutions) should be sent to the appropriate one of the associate editors, as follows:

Notes: CAROL G. CRAWFORD, Department of Mathematics, U.S. Naval Academy, Annapolis, MD 21402, *or* RICHARD LIBERA, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716, *or* ANITA E. SOLOW, Department of Mathematics, Grinnell College, Grinnell, IA 50112.

Teaching of Mathematics: JOAN P. HUTCHINSON *or* STAN WAGON, Department of Mathematics, Smith College, Northampton, MA 01063.

Unsolved Problems: RICHARD GUY, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.

Proposed problems and solutions, both elementary and advanced: PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801.

Reprint Permission: A. B. WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Back Issues: P. and H. BLISS Co., Middletown, CT 06457.

Microfilm Editions: University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

Editorial Assistants: Fanny Hailpern, Mary P. Jackson.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT, ten times a year: January, February, March, April, May, June-July, August-September, October, November and December.

The annual subscription price for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is $24 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $24. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $70 per year.

Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

PRINTED IN THE UNITED STATES OF AMERICA

SHUFFLING CARDS AND STOPPING TIMES

DAVID ALDOUS*
Department of Statistics, University of California, Berkeley, CA 94720

PERSI DIACONIS**
Department of Statistics, Stanford University, Stanford, CA 94305

1. Introduction. How many times must a deck of cards be shuffled until it is close to random? There is an elementary technique which often yields sharp estimates in such problems. The method is best understood through a simple example.

EXAMPLE 1. *Top in at random shuffle.* Consider the following method of mixing a deck of cards: the top card is removed and inserted into the deck at a random position. This procedure is repeated a number of times. The following argument should convince the reader that about $n \log n$ shuffles suffice to mix up n cards. The argument depends on following the bottom card of the deck. This card stays at the bottom until the first time (T_1) a card is inserted below it. Standard calculations, reviewed below, imply this takes about n shuffles. As the shuffles continue, eventually a second card is inserted below the original bottom card (this takes about $n/2$ further shuffles). Consider the instant (T_2) that a second card is inserted below the original bottom card. The two cards under the original bottom card are equally likely to be in relative order low-high or high-low.

Similarly, the first time a third card is inserted below the original bottom card, each of the 6 possible orders of the 3 bottom cards is equally likely. Now consider the first time T_{n-1} that the original bottom card comes up to the top. By an inductive argument, all $(n-1)!$ arrangements of the lower cards are equally likely. When the original bottom card is inserted at random, at time $T = T_{n-1} + 1$, then all $n!$ possible arrangements of the deck are equally likely.

	1	2	3	4	5	6	7	8	9
a	*b*	*c*	*c*	*a*	*b*	*a*	*a*	*d*	*c*
b	*c*	*a*	*a*	*b*	*a*	*d*	*d*	*c*	*d*
c	*a*	*b*	*b*	*d*	*d*	*c*	*c*	*a*	*a*
d	*d*	*d*	*d*	*c*	*c*	*b*	*b*	*b*	*b*
				T_1		T_2		T_3	T

FIG. 1. Example of repeated top in at random shuffles of a 4-card deck.

When the original bottom card is at position k from the bottom, the waiting time for a new card to be inserted below it is about n/k. Thus the waiting time T for the bottom card to come to

David Aldous confesses to a conventional career, going from a Ph.D. and Research Fellowship at Cambridge University to the Statistics Department at Berkeley, where he is now Associate Professor. He does research in theoretical and applied probability theory, and for recreation he plays volleyball (well), bridge (badly) and watches Monty Python reruns.

Persi Diaconis left High School at an early age to earn a living as a magician and gambler, only later to become interested in mathematics and earn a Ph.D. at Harvard. After a spell at Bell Labs, he is now Professor in the Statistics Department at Stanford. He was an early recipient of a MacArthur Foundation award, and his wide range of mathematical interests is partly reflected in his first book *Group Theory in Statistics.* He retains an interest in magic and the exposure of fraudulent psychics.

*Research supported by National Science Foundation Grant MCS80-02698.
**Research supported by National Science Foundation Grant MCS80-24649.

the top and be inserted is about

$$n + \frac{n}{2} + \frac{n}{3} + \cdots + \frac{n}{n} \doteq n \log n.$$

This paper presents a rigorous version of the argument and illustrates its use in a variety of random walk problems. The next section introduces the basic mathematical set up. Section 3 details a number of examples drawn from applications such as computer generated pseudo random numbers. Section 4 treats ordinary riffle shuffling, analyzing a model introduced by Gilbert, Shannon, and Reeds. Section 5 explains a sense in which the method of stopping times always works and compares this to two other techniques (Fourier analysis and coupling). Some open problems are listed.

2. The Basic Set-Up. Repeated shuffling is best treated as random walk on the permutation group S_n. For later applications, we treat an arbitrary finite group G. Given some scheme for randomly picking elements of G, let $Q(g)$ be the probability that g is picked. The numbers $\{Q(g): g \in G\}$ are a (probability) *distribution*: $Q(g) \geqslant 0$ and $\Sigma Q(g) = 1$. Repeated independent picks according to the same scheme yield random elements $\xi_1, \xi_2, \xi_3, \ldots,$ of G. Define the products

$$\begin{aligned} X_0 &= \text{identity} \\ X_1 &= \xi_1 \\ &\vdots \\ X_k &= \xi_k X_{k-1} = \xi_k \xi_{k-1} \cdots \xi_1. \end{aligned}$$

The random variables $X_0, X_1, X_2, \ldots,$ are the *random walk* on G with *step distribution* Q. Think of X_k as the position at time k of a randomly-moving particle. The distribution of X_2, that is the set of probabilities $P(X_2 = g)$, $g \in G$, is given by convolution

$$P(X_2 = g) = Q^*Q(g) = \sum_{h \in G} Q(h) Q(gh^{-1}).$$

For $Q(h)Q(gh^{-1})$ is the chance that element h was picked first and gh^{-1} was picked second; for any h, this makes the product equal to g. Similarly, $P(X_k = g) = Q^{k*}(g)$, where Q^{k*} is the repeated convolution

$$Q^{k*} = Q^*Q^{(k-1)*} = \sum_{h \in G} Q(h) Q^{(k-1)*}(gh^{-1}). \tag{2.1}$$

In modelling shuffling of an n-card deck, the state of the deck is represented as a permutation $\pi \in S_n$, meaning that the card originally at position i is now at position $\pi(i)$.

In Example 1, $G = S_n$, and using cycle notation for permutations π,

$$\begin{aligned} Q(i, i-1, \ldots, 1) &= 1/n, \quad 1 \leqslant i \leqslant n, \\ Q(\pi) &= 0, \text{else}. \end{aligned}$$

Here ξ_k is a randomly chosen cycle, X_k is the state of the deck after k shuffles, and $Q^{k*}(\pi)$ is the chance that the state after k shuffles is π. In Fig. 1, $\xi_1 = (3,2,1)$, $\xi_2 = (3,2,1)$ and $X_2 = \xi_1^* \xi_2 = (1,2,3)$.

We shall study the distribution Q^{k*}. Note that Q^{k*} can be defined by (2.1) without using the richer structure of the random walk (X_k); however, this richer structure is essential for our method of analysis.

A fundamental result is that repeated convolutions converge to the uniform distribution U:

$$Q^{k*}(g) \to U(g) = 1/|G| \quad \text{as } k \to \infty, \tag{2.2}$$

unless Q is concentrated on some coset of some subgroup. This was first proved by Markov (1906)—see Feller (1968), Section 15.10 for a clear discussion—and can nowadays be regarded as a special case of the basic limit theory of finite Markov chains. Poincaré (1912) gave a Fourier

analytic proof, and subsequent workers have extended (2.2) to general compact groups—see Grenander (1963), Heyer (1977), Diaconis (1982) for surveys. A version of this result is given here as Theorem 3 of Section 3.

Despite this work on abstracting the asymptotic result (2.2), little attention has been paid until recently to the kind of non-asymptotic questions which are the subject of this paper.

A natural way to measure the difference between two probability distributions Q_1, Q_2 on G is by *variation distance*

$$\|Q_1 - Q_2\| = \frac{1}{2}\sum|Q_1(g) - Q_2(g)|.$$

There are equivalent definitions

$$\|Q_1 - Q_2\| = \max_{A\subset G}|Q_1(A) - Q_2(A)| = \frac{1}{2}\max_{\|f\|=1}|Q_1(f) - Q_2(f)|, \tag{2.3}$$

where $Q(A) = \Sigma_{g\in A}Q(g)$, $Q(f) = \Sigma f(g)Q(g)$, and $\|f\| = \max|f(g)|$. The string of equalities is proved by noting that the maxima occur for $A = \{g : Q_1(g) > Q_2(g)\}$ and for $f = 1_A - 1_{\bar{A}}$. Thus, two distributions are close in variation distance if and only if they are uniformly close on all subsets. Plainly $0 \leqslant \|Q_1 - Q_2\| \leqslant 1$.

An example may be useful. Suppose, after well-shuffling a deck of n cards, that you happen to see the bottom card, c. Then your distribution Q on S_n is uniform on the set of permutations π for which $\pi(c) = n$, and $\|Q - U\| = 1 - 1/n$. This shows the variation distance can be very "unforgiving" of small deviations from uniformity.

Given a distribution Q on a group G, (2.2) says

$$d_Q(k) \overset{\text{def}}{\equiv} \|Q^{k*} - U\| \to 0 \quad \text{as } k \to \infty. \tag{2.4}$$

Where Q models a random shuffle, $d(k)$ measures how close k repeated shuffles get the deck to being perfectly (uniformly) shuffled. One might suppose $d(k)$ decreases smoothly from (near) 1 to 0; and it is not hard to show $d(k)$ is decreasing. However,

THEOREM 1. *For the "top in at random" shuffle*, Example 1,
(a) $d(n \log n + cn) \leqslant e^{-c}$; *all* $c \geqslant 0$, $n \geqslant 2$.
(b) $d(n \log n - c_n n) \to 1$ *as* $n \to \infty$; *all* $c_n \to \infty$.

This gives a sharp sense to the assertion that $n \log n$ shuffles are enough. This is a particular case of a general *cut-off phenomenon*, which occurs in all shuffling models we have been able to analyze; there is a critical number k_0 of shuffles such that $d(k_0 + o(k_0)) \approx 0$ but $d(k_0 - o(k_0)) \approx 1$. (See Fig. 2.)

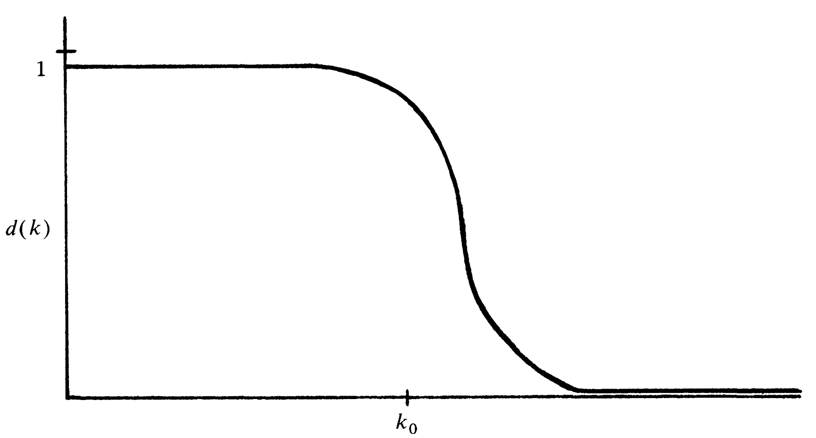


FIG. 2

Our aim is to determine k_0 in particular cases. This is quite different from looking at asymptotics in (2.4): it is elementary that $d(k) \to 0$ geometrically fast, and Perron-Frobenius theory says $d(k) \sim a\lambda^k$, where a, λ have eigenvalue/eigenvector interpretation, but these asymptotics miss the cut-off phenomenon. For card players, the question is not "exactly how close to uniform is the deck after a million riffle shuffles?", but "is 7 shuffles enough?".

The main purpose of this paper is to show how upper bounds on $d(k)$, like (a) in Theorem 1, can be obtained using the notion of strong uniform times, which we now define in two steps.

DEFINITION 1. Let G be a finite group, and G^∞ the set of all G-valued infinite sequences $\mathbf{g} = (g_1, g_2, \dots)$. A *stopping rule* $\hat{T}$ is a function $\hat{T}$: $G^\infty \to \{1, 2, 3, \dots; \infty\}$ such that if $\hat{T}(\mathbf{g}) = j$, then $\hat{T}(\hat{\mathbf{g}}) = j$ for all $\hat{\mathbf{g}}$ with $\hat{g}_i = g_i$, $i \leqslant j$.

DEFINITION 2. Let Q be a distribution on G, and let (X_k) be the associated random walk. Given a stopping rule $\hat{T}$, the random time $T = \hat{T}(X_1, X_2, \dots)$ is a *stopping time*. Call T a strong uniform time (for U) if for each $k < \infty$

(a) $P(T = k,\ X_k = g)$ does not depend on g.

REMARK (i). Note that (a) is equivalent to

(b) $P(X_k = g | T = k) = 1/|G|$, $g \in G$

and to

(c) $P(X_k = g | T \leqslant k) = 1/|G|$; $g \in G$.

REMARK (ii). Picture the process of picking group elements and multiplying. A stopping time is a rule which tells you when to "stop" with the current value of the product. The time is strong uniform if, conditional on stopping after k steps, the value of the product is uniform on G.

REMARK (iii). In Example 1, we defined a time T as the first time that the original bottom card has come to the top and been inserted into the deck. This is certainly a stopping time, and the inductive argument in Section 1 shows that, given $T = k$, all arrangements of the deck are equally likely.

REMARK (iv). In practice it is often useful to have a slightly more general notion of stopping time, which allows the decision on whether or not to stop at n to depend not only on $(X_1, \dots, X_n)$ but also on the value of some random quantity Y independent of the X process. Such a time T is called a *randomized* stopping time T; our results extend to this case without essential change.

Here is a basic upper bound lemma which relates strong uniform times to the distance between Q^{k*} and the uniform distribution.

LEMMA 1. *Let Q be a probability distribution on a finite group G. Let T be a strong uniform time for Q. Then*

$$d(k) \equiv \|Q^{k*} - U\| \leqslant P(T > k), \quad \text{all } k \geqslant 0.$$

Proof. For any $A \subset G$

$$\begin{aligned} Q^{k*}(A) &= P(X_k \in A) \\ &= \sum_{j \leqslant k} P(X_k \in A, T = j) + P(X_k \in A, T > k) \\ &= \sum_{j \leqslant k} U(A) P(T = j) + P(X_k \in A | T > k) P(T > k) \\ &= U(A) + \{P(X_k \in A | T > k) - U(A)\} P(T > k) \end{aligned}$$

and so

$$|Q^{k*}(A) - U(A)| \leqslant P(T > k). \quad \square$$

We conclude this section by using Lemma 1 and elementary probability concepts to prove Theorem 1. Here is one elementary result we shall use in several examples.

LEMMA 2. *Sample uniformly with replacement from an urn with n balls. Let V be the number of draws required until each ball has been drawn at least once. Then*

$$P(V > n\log n + cn) \leqslant e^{-c}; \quad c \geqslant 0, n \geqslant 1.$$

Proof. Let $m = n\log n + cn$. For each ball b let A_b be the event "ball b not drawn in the first m draws". Then

$$P(V > m) = P\Big(\bigcup_b A_b\Big) \leqslant \sum_b P(A_b) = n\left(1 - \frac{1}{n}\right)^m$$
$$\leqslant n\exp(-m/n) = e^{-c}. \quad \square$$

REMARK. This is the famous "coupon-collector's problem", discussed in Feller (1968). The asymptotics are $P(V > n\log n + cn) \to 1 - \exp(-e^{-c})$ as $n \to \infty$, c fixed. So for c not small the bound in Lemma 2 is close to sharp.

Proof of Theorem 1. Recall we have argued that T, the first time that the original bottom card has come to the top and been inserted into the deck, is a strong uniform time for this shuffling scheme. We shall prove that T has the same distribution as V in Lemma 2; then assertion (a) is a consequence of Lemmas 1 and 2.

We can write

$$T = T_1 + (T_2 - T_1) + \cdots + (T_{n-1} - T_{n-2}) + (T - T_{n-1}), \tag{2.5}$$

where T_i is the time until the ith card is placed under the original bottom card. When exactly i cards are under the original bottom card b, the chance that the current top card is inserted below b is $\dfrac{i+1}{n}$, and hence the random variable $T_{i+1} - T_i$ has geometric distribution

$$P(T_{i+1} - T_i = j) = \frac{i+1}{n}\left(1 - \frac{i+1}{n}\right)^{j-1}; \quad j \geqslant 1. \tag{2.6}$$

The random variable V in Lemma 2 can be written as

$$V = (V - V_{n-1}) + (V_{n-1} - V_{n-2}) + \cdots + (V_2 - V_1) + V_1, \tag{2.7}$$

where V_i is the number of draws required until i distinct balls have been drawn at least once. After i distinct balls have been drawn, the chance that a draw produces a not-previously-drawn ball is $\dfrac{n-i}{n}$. So $V_i - V_{i-1}$ has distribution

$$P(V_i - V_{i-1} = j) = \frac{n-i}{n}\left(1 - \frac{n-i}{n}\right)^{j-1}; \quad j \geqslant 1.$$

Comparing with (2.6), we see that corresponding terms $(T_{i+1} - T_i)$ and $(V_{n-i} - V_{n-i-1})$ have the same distribution; since the summands within each of (2.5) and (2.7) are independent, it follows that the sums T and V have the same distribution, as required.

To prove (b), fix j and let A_j be the set of configurations of the deck such that the bottom j original cards remain in their original relative order. Plainly $U(A_j) = 1/j!$ Let $k = k(n)$ be of the form $n\log n - c_n n$, $c_n \to \infty$. We shall show

$$Q^{k*}(A_j) \to 1 \quad \text{as } n \to \infty; \quad j \text{ fixed}. \tag{2.8}$$

Then $d(k) \geqslant \max_j\{Q^{k*}(A_j) - U(A_j)\} \to 1$ as $n \to \infty$, establishing part (b).

To prove (2.8), observe that $Q^{k*}(A_j) \geqslant P(T - T_{j-1} > k)$. For $T - T_{j-1}$ is distributed as the

time for the card initially jth from bottom to come to the top and be inserted; and if this has not occurred by time k, then the original bottom j cards must still be in their original relative order at time k. Thus it suffices to show

$$P(T - T_{j-1} \leqslant k) \to 0 \quad \text{as } n \to \infty; \quad j \text{ fixed.} \tag{2.9}$$

We shall prove this using *Chebyshev's inequality*:

$$P(|Z - EZ| \geqslant a) \leqslant \frac{\operatorname{var}(Z)}{a^2}, \quad \text{where } a \geqslant 0, \text{ and } Z \text{ is any random variable.}$$

From (2.6),

$$E(T_{i+1} - T_i) = \frac{n}{i+1}, \quad \operatorname{var}(T_{i+1} - T_i) = \left(\frac{n}{i+1}\right)^2 \left(1 - \frac{i+1}{n}\right),$$

and so from (2.5)

$$E(T - T_j) = \sum_{i=j}^{n-1} \frac{n}{i+1} = n \log n + O(n),$$

$$\operatorname{var}(T - T_j) = \sum_{i=j}^{n-1} \left(\frac{n}{i+1}\right)^2 \left(1 - \frac{i+1}{n}\right) = O(n^2),$$

and Chebyshev's inequality applied to $Z = T - T_{j-1}$ readily yields (2.9). □

REMARK. Note that the "strong uniform time" property of T played no role in establishing the *lower* bound (b). Essentially, we get lower bounds by guessing some set A for which $|Q^{k*}(A) - U(A)|$ should be large, and using the obvious (from (2.3)) inequality

$$d(k) = \|Q^{k*} - U\| \geqslant |Q^{k*}(A) - U(A)|.$$

3. Examples. We present constructions of strong uniform times for a variety of random walks: simple random walk on the circle, general random walks on finite groups, and a random walk arising in random number generation. Sometimes our arguments give the optimal rate, often they give the correct order of magnitude.

EXAMPLE 2. *Simple random walk on the integers* mod n. Let n be a positive odd integer. Let Z_n be the integers mod n, thought of as n points on a circle. Imagine a particle which moves by steps, each step being equally likely to move 1 to the right or 1 to the left. This random walk has step distribution Q on Z_n;

$$Q(1) = Q(-1) = \frac{1}{2}. \tag{3.1}$$

The following theorem shows that the number of steps k required to become uniform is slightly more than n^2.

THEOREM 2. *Let $n \geqslant 3$ be an odd integer. For simple random walk on the integers* mod n *defined by* (3.1), *for $k \geqslant n^2$,*

$$d(k) \leqslant 6e^{-ak/n^2}$$

with $a = 4\pi^2/3$.

Proof. First consider $n = 5$ and the following 5 patterns

$$++--, \; +---, \; -+++, \; ++++, \; ----.$$

A sequence of successive steps of the walk on Z_5 yields a sequence of $\pm$ symbols. Consider the sequence in disjoint blocks of 4. Stop the first time T that a block of 4 equals one of the above 5 patterns. Thus, if the sequence starts $++-+, +++-, ++--$, $T = 12$.

This stopping time is clearly a strong uniform time; given that $T = 12$, all 5 final positions in Z_5 are equally likely. Such sets of k-tuples can be chosen for any odd n. It turns out that to get the correct rate of convergence, k should be chosen as a large multiple of n^2. Here are some details.

For fixed integers n and k, with n odd, let B_j be the set of binary k-tuples with j pluses (mod n).

Let j^* be the index corresponding to the smallest $|B_{j^*}|$. Partition the set of binary k-tuples into n groups of size $|B_{j^*}|$, the jth group being chosen arbitrarily from B_j. The random walk generates a sequence of $\pm$ symbols. Consider these in disjoint blocks of length k. Define T as the first time a block equals one of the chosen group. This clearly yields a strong uniform time. The following lemma gives an explicit upper bound for $d(k)$.

LEMMA 3. *Let T be as defined above. For $n \geqslant 3$ and $k \geqslant n^2$,*

$$P(T > k) \leqslant 6e^{-ak/n^2}$$

with $a = 4\pi^2/3$.

Proof. The number of elements in B_j is

$$\sum_{l \geqslant 0} \binom{k}{ln + j} = \frac{2^k}{n} \sum_{l=0}^{n-1} e^{\frac{-2\pi i l j^*}{n}} \left(\cos \frac{2\pi l}{n} \right)^k,$$

this being a classical identity due to C. Ramus (see Knuth (1973, p. 70)). The chance of a given block falling in the chosen group equals

$$p \stackrel{\text{def}}{=} \frac{n}{2^k} |B_{j^*}| = \sum_{l=0}^{n-1} e^{\frac{-2\pi i l j}{n}} \left(\cos \frac{2\pi l}{n} \right)^k.$$

Now

$$P(T > k) = p(1-p) + p(1-p)^2 + p(1-p)^3 + \cdots = 1 - p \leqslant \sum_{l=1}^{n-1} \left| \cos \frac{2\pi l}{n} \right|^k.$$

Straightforward calculus using quadratic approximations to cosine such as $\cos x \leqslant 1 - \dfrac{x^2}{3} \leqslant e^{-x^2/3}$ for $0 \leqslant x \leqslant \pi/2$ leads to the stated result. Further details may be found in Chung, Diaconis, and Graham (1986). □

REMARK. There is a lower bound for $d(k)$ of the form $\alpha e^{-\beta k/n^2}$ for positive α and β, so somewhat more than n^2 steps really are required. One way to prove this is to use the central limit theorem; this implies that after k steps the walk has moved a net distance of order $k^{1/2}$. Hence we need k of order n^2 at least in order that the distribution after k steps is close to uniform. Further details are in Chung, Diaconis and Graham (1986).

There is a sense in which the cutoff phenomenon does not occur for this example. It is possible to show there is a continuous function $d^*(t)$, with $d^*(t) \to 0$ as $t \to \infty$, such that for simple random walk on Z_n,

$$\max_k |d(k) - d^*(k/n^2)| \to 0 \quad \text{as } n \to \infty.$$

Indeed, as $n \to \infty$, a rescaled version of the random walk tends to Brownian motion on the circle. The function $d^*(t)$ is the variation distance to uniformity for Brownian motion at time t.

EXAMPLE 3. *A bound for general problems.* Let G be a finite group and Q a probability on G. The following result shows that Q^{*k} converges to the uniform distribution geometrically fast provided Q is not concentrated on a subgroup or a translate of a subgroup. To see the need for this condition, consider Example 2 above (simple random walk on Z_n). If n is even, then the

particle is at an even position after an even number of steps—the distribution never converges to uniform.

A simple way to force convergence is the following:

(3.2) Suppose for some k_0 and $0 < c < 1$, $Q^{*k_0}(g) \geqslant cU(g)$ for all $g \in G$.

THEOREM 3. *Condition* (3.2) *implies*

$$d(k) \leqslant (1-c)^{[k/k_0]} \quad \text{for all } k \geqslant k_0.$$

Proof. The argument proceeds by constructing another process which behaves like the original random walk but easily exhibits a strong uniform time. Suppose first that $k_0 = 1$, so $Q(g) \geqslant cU(g)$ for all g. Define

$$R(g) = [Q(g) - cU(g)]/[1-c].$$

Observe that $R(g)$ is a probability and

$$Q(g) = (1-c)R(g) + cU(g). \tag{3.3}$$

Consider a new random walk defined as follows. For each step, flip a coin with probability of heads c. If the coin comes up heads, take the step according to $U(g)$. If the coin comes up tails, take the step according to $R(g)$. Because of (3.3), each step is taken according to Q overall. Let T be the first time that the coin comes up heads. Then T is a (randomized) stopping time and because the convolution of the uniform distribution with any distribution is uniform, T is a strong uniform time.

Clearly,

$$P\{T > k\} = (1-c)^k.$$

For $k_0 > 1$, apply the argument to the probability Q^{*k_0}. □

REMARK (i). The argument given is valid for a probability on a general compact group. In this form, Theorem 3 is due to Kloss (1959). The proof we give is very close to techniques exploited by Athreya and Ney (1977) for general state space Markov processes.

REMARK (ii). While Theorem 3 seems quantitative, the simplicity of the argument should make one suspicious. The reader can see the difficulty by trying to get a rate of convergence for simple random walk on Z_n. Estimating c and k_0 is not an easy problem, we do not know how to use Theorem 3 to get the correct rate of convergence for any non-trivial problem.

EXAMPLE 4. *A random walk on Z_n arising in random number generation.* Random number generators often work by recursively computing $X_{k+1} = aX_k + b$ (modulo n), where a, b and n are chosen carefully—see Knuth (1981). Of course the sequence X_k is really deterministic and exhibits many regularities. To improve things, various schemes have been suggested involving combining several generators. In one scheme, a and b are chosen each step from another generator. If this second source is considered truly random (it may be the result of a physical generator using a radioactive source) one may inquire how long it takes X_k to become random.

For example, if $a = 1$ and $b = 0$, $+1$, or -1 each with probability $1/3$, the process becomes simple random walk on Z_n: $X_k = X_{k-1} + b_k \pmod n$ with a slightly different step size than considered in Example 2. The argument given there can easily be adapted to show that slightly more than n^2 steps are required to become random.

We now consider the effect of a deterministic doubling:

$$X_k = 2X_{k-1} + b_k \pmod n, \quad b_k = 0, \pm 1 \text{ with probability } \frac{1}{3}. \tag{3.4}$$

We will show that this dramatically speeds things up: from n^2 down to $\log n \log\log n$. The argument is presented as a non-trivial illustration of the method of strong uniform times. It

involves a novel construction of an almost uniform time. For simplicity, we take $n = 2^l - 1$ (a common choice in the application).

THEOREM 4. *Let Q_k be the probability distribution of X_k defined by* (3.4) *with $n = 2^l - 1$. Let $d(k) = \|Q_k - U\|$. Then*

$$d(cl\log l) \to 0 \quad \text{as } l \to \infty, \quad \text{for } c > \frac{1}{\log 3}.$$

Proof. Observe first that if δ_i takes values ± 1 with probability $1/2$, then

$$U^* = 2^{l-1}\delta_1 + 2^{l-2}\delta_2 + \cdots + \delta_l$$

is very close to uniformly distributed $\bmod 2^l - 1$. Indeed,

$$P\big(U^* = j(\bmod 2^l - 1)\big) = \begin{cases} \dfrac{1}{2^l}, & j \neq 0, \\ \dfrac{2}{2^l}, & j = 0. \end{cases}$$

Thus

$$\|U^* - U\| = \frac{2}{2^{l-1}} - \frac{1}{2^l - 1}. \tag{3.5}$$

The argument proceeds by finding a stopping time T such that the process stopped at time T has distribution at least as close to uniform as U^*. An appropriate modification of the upper bound lemma will complete the proof. We isolate the steps as a sequence of lemmas. The first and second lemmas are elementary with proofs omitted.

LEMMA 4. *Let $X_1, X_2, \ldots$ be a process with values in a finite group G. Write Q_k for the probability distribution of X_k. Let T be a stopping time with the property that for some $\varepsilon > 0$,*

$$\|Q_k(\cdot | T = j) - U\| \leqslant \varepsilon; \quad \text{all } j \leqslant k.$$

Then

$$\|Q_k - U\| \leqslant \varepsilon + P(T > k).$$

LEMMA 5. *Let Q_1 and Q_2 be probability distributions on a finite group G. Then*

$$\|Q_1^* Q_2 - U\| \leqslant \|Q_1 - U\|.$$

To state the third lemma, an appropriate stopping time T must be defined. Using the defining recurrence $X_k = 2X_{k-1} + b_k(\bmod n)$,

$$X_k = 2^{k-1}b_1 + 2^{k-2}b_2 + \cdots + b_k(\bmod n). \tag{3.6}$$

Since $n = 2^l - 1$, $2^l = 1(\bmod n)$. Group the terms on the right side of (3.6) by distinct powers of 2:

$$X_k = 2^{l-1}A_1 + 2^{l-2}A_2 + \cdots + A_l(\bmod n)$$

with

$$A_1 = b_1 + b_{l+1} + b_{2l+1} \ldots, \; A_2 = b_2 + b_{l+2} + \cdots, \text{etc.}$$

Define T as the first time each of the sums $A_1, A_2, \ldots, A_l$ contains at least one non-zero summand.

LEMMA 6. *The probability distribution of X_k given $T = j < k$ is the convolution of U^* defined above with an independent random variable.*

Proof. Let δ_i^* be the first non-zero summand in A_i. Write

$$X_k = \left[2^{l-1}\delta_1^* + 2^{l-2}\delta_2^* + \cdots + \delta_l^*\right] + \left[2^{l-1}(A_1 - \delta_1^*) + \cdots + (A_l - \delta_l^*)\right].$$

Clearly the first term on the right has distribution U^*. Further,, given all the remaining values of b_k, and the labels of δ_i^*, all 2^l values of $\delta_1^*, \ldots, \delta_l^*$ are equally likely, so the decomposition of X_k is into independent parts. □

Using Lemmas 5, 6, along with the bound (3.5) allows us to take $\varepsilon = 2/2^l$ in Lemma 4 for this stopping time T. To complete the proof of Theorem 4, it only remains to estimate $P(T > k)$.

Toward this end, consider $k = al$ for integer a,

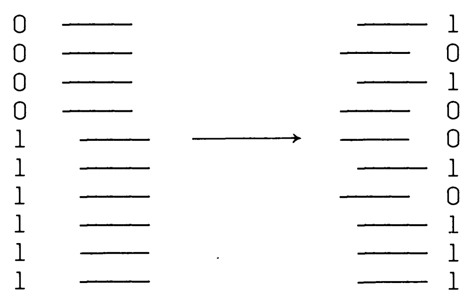
$$P(T > al) = 1 - \left(1 - \left(\frac{1}{3}\right)^a\right)^l.$$

For large l, this is approximately $1 - \exp\{-le^{-a\log 3}\}$. If $a = \dfrac{\log l + c}{\log 3}$ for some value of c, this becomes $1 - \exp\{-e^{-c}\}$ which is well approximated by e^{-c} for large c. It follows that for c large, $\dfrac{l\log l}{\log 3} + cl$ steps suffice to be close to uniform. This is more than was claimed in Theorem 4. □

REMARK. Chung, Diaconis and Graham (1986) give a more detailed analysis, showing that $l\log l$ is the correct order of magnitude.

4. An Analysis of Riffle Shuffles. In this section we analyze the most commonly used method of shuffling cards—the ordinary riffle shuffle. This involves cutting the deck approximately in half, and interleaving (or riffling) the two halves together. We begin by introducing a mathematical model for shuffling suggested by Gilbert, Shannon and Reeds. Following Reeds, we introduce a strong uniform time for this model and show how the calculations reduce to simple facts about the birthday problem.

The diagram gives the result of a single riffle shuffle of a 10 card deck in the usual $i \to \pi(i)$ format



i	π(i)
1	2
2	4
3	5
4	7
5	1
6	3
7	6
8	8
9	9
10	10

This shuffle is the result of cutting 4 cards off the top of a 10 card "deck" and riffling the packets together, first dropping cards 10, 9, 8, then card 4, then 7, and so on.

This permutation has two rising sequences

$$\pi(1) < \pi(2) < \pi(3) < \pi(4) \quad \text{and} \quad \pi(5) < \pi(6) < \pi(7) < \pi(8) < \pi(9) < \pi(10).$$

In general, a permutation π of n cards made by a riffle shuffle will have exactly 2 rising sequences (unless it is the identity, which has 1). Conversely, any permutation of n cards with 1 or 2 rising sequences can be obtained by a physical riffle. Thus the mathematical definition of *a riffle shuffle*

is "a permutation with 1 or 2 rising sequences". Suppose c cards are initially cut off the top. Then there are $\binom{n}{c}$ possible riffle shuffles (1 of which is the identity). To see why, label each of the c cards cut with "0" and the others with "1". After the shuffle, the labels form a binary n-tuple with c "0"s: there are $\binom{n}{c}$ such n-tuples and each corresponds to a unique riffle shuffle. Finally, the total number of possible riffle shuffles is

$$1 + \sum_{c=0}^{n} \left\{ \binom{n}{c} - 1 \right\} = 2^n - n.$$

Some stage magicians can perform "perfect" shuffles, but for most of us the result of a shuffle is somewhat random. The actual distribution of one shuffle (that is, the set of probabilities of each of the $2^n - n$ possible riffle shuffles) will depend on the skill of the individual shuffler. The following model for random riffle shuffle, suggested by Gilbert and Shannon (1955) and Reeds (1981), is mathematically tractable and qualitatively similar to shuffles done by amateur card players.

1st description. Begin by choosing an integer c from $0, 1, \ldots, n$ according to the binomial distribution $P\{C = c\} = \frac{1}{2^n}\binom{n}{c}$. Then, c cards are cut off and held in the left hand, and $n - c$ cards are held in the right hand. The cards are dropped from a given hand with probability proportional to packet size. Thus, the chance that a card is first dropped from the left hand packet is c/n. If this happens, the chance that the next card is dropped from the left packet is $(c-1)/(n-1)$.

There are two other descriptions of this shuffling mechanism that are useful.

2nd description. Cut an n card deck according to a binomial distribution. If c cards are cut off, pick one of the $\binom{n}{c}$ possible shuffles uniformly.

3rd description. This generates π^{-1} with the correct probability. Label the back of each card with the result of an independent, fair coin flip as 0 or 1. Remove all cards labelled 0 and place them on top of the deck, keeping them in the same relative order.

LEMMA 7. *The three descriptions yield the same probability distribution.*

Proof. The second and third descriptions are equivalent. Indeed, the binary labelling chooses a binomially distributed number of zeros, and conditional on this choice, all possible placements of the zeros are equally likely.

The first and second descriptions are equivalent. Suppose c cards have been cut off. For the first description, a given shuffle is specified by a sequence $D_1, D_2, \ldots, D_n$, where each D_i can be L or R and c of the D_i's must be L. Under the given model, the chance of all such sequences, determined by multiplying the chance at each stage, is $c!(n-c)!/n!$ □

The argument to follow analyzes the repeated inverse shuffle. This has the same distance to uniform as repeated shuffling because of the following lemma.

LEMMA 8. *Let G be a finite group, $T: G \to G$ one-to-one, and Q a probability on G. Then*

$$\|Q - U\| = \|QT^{-1} - U\|,$$

where $QT^{-1}(g) = Q(T^{-1}(g))$ is the probability induced by T. □

The results of repeated inverse shuffles of n cards can be recorded by forming a binary matrix with n rows. The first column records the zeros and ones that determine the first shuffle, and so on. The ith row of the matrix is associated to the ith card in the original ordering of the deck, recording in coordinate j the behavior of this card on the jth shuffle.

	1	2	3	4
a 1101	c 0010	c 0010	f 1000	f 1000
b 1100	e 0110	d 1011	a 1101	b 1100
c 0010	a 1101	f 1000	b 1100	c 0010
d 1011	b 1100	e 0110	c 0010	e 0110
e 0110	d 1011	a 1101	d 1011	a 1101
f 1000	f 1000	b 1100	e 0110	d 1011

LEMMA 9 (Reeds). *Let T be the first time that the binary matrix formed from inverse shuffling has distinct rows. Then T is a strong uniform time.*

Proof. The matrix can be considered as formed by flipping a fair coin to fill out the i, j entry. At every stage, the rows are independent binary vectors. The joint distribution of the rows, conditional on being all distinct, is invariant under permutations.

After the first inverse shuffle, all cards associated to binary vectors starting with 0 are above cards with binary vectors starting with 1. After two shuffles, cards associated with binary vectors starting $(0,0)$ are on top followed by cards associated to vectors beginning $(1,0)$, followed by $(0,1)$, followed by $(1,1)$ at the bottom of the deck.

Inductively, the inverse shuffles sort the binary vectors (from right to left) in lexicographic order. At time T the vectors are all distinct, and all sorted. By permutation invariance, any of the n cards is equally likely to have been associated with the smallest row of the matrix (and so be on top). Similarly, at time T, all $n!$ orders are equally likely. □

To complete this analysis, the chance that $T > k$ must be computed. This is simply the probability that if n balls are dropped into 2^k boxes there are not two or more balls in a box. If the balls are thought of as people, and the boxes as birthdays, we have the familiar question of the birthday problem and its well-known answer. This yields:

THEOREM 5. *For Q the Gilbert-Shannon-Reeds distribution defined in Lemma 7,*

$$\|Q^{*k} - U\| \leqslant P(T > k) = 1 - \prod_{i=1}^{n-1}\left(1 - \frac{i}{2^k}\right). \tag{4.1}$$

Standard calculus shows that if $k = 2\log_2(n/c)$,

$$P(T > k) \overset{n\to\infty}{\sim} 1 - e^{-\frac{c^2}{2}} \overset{c\to 0}{\sim} \frac{1}{2}c^2.$$

In this sense, $2\log_2 n$ is the cut off point for this bound. Exact computation of the right side of (4.1) when $n = 52$ gives the bounds

k	upper bound
10	.73
11	.48
12	.28
13	.15
14	.08

REMARK (a). The lovely new idea here is to consider shuffling as inverse sorting. The argument works for any symmetric method of labelling the cards. For example, biased cuts can be modeled by flipping an unfair coin. To model cutting off exactly j cards each time, fill the columns of the matrix with the results of n draws without replacement from an urn containing j balls labelled zero and $n - j$ balls labelled one. These lead to slightly unorthodox birthday problems which turn out to be easy to work with.

REMARK (b). The argument can be refined. Suppose shuffling is stopped slightly before all rows of the matrix are distinct—e.g., stop after $2\log n$ shuffles. Cards associated to identical binary rows correspond to cards in their original relative positions. It is possible to bound how far such permutations are from uniform and get bounds on $\|Q^{*k} - U\|$. Reeds (1981) has used such arguments to show that 9 or fewer shuffles make the variation distance small for 52 cards.

REMARK (c). A variety of ad hoc techniques have been used to get lower bounds. One simple method that works well is to simply follow the top card after repeated shuffles. This executes a Markov chain on n states with a simple transition matrix. For n in the range of real deck sizes, $n \times n$ matrices can be numerically multiplied and then the variation distance to uniform computed. Reeds (1981) has carried this out for decks of size 52 and shown that $\|Q^{*6} - U\| \geqslant .1$. Techniques which allow asymptotic verification that $k = (3/2)\log_2 n$ is the right cutoff for large n are described in Aldous (1983a). These analyses, and the results quoted above, suggest that seven riffle shuffles are needed to get close to random.

REMARK (d). Other mathematical models for riffle shuffling are suggested in Donner and Uppulini (1970), Epstein (1977), and Thorp (1973). Borel and Cheron (1955) and Kosambi and Rao (1958) discuss the problem in a less formal way. Where conclusions are drawn, 6 to 7 shuffles are recommended to randomize 52 cards.

REMARK (e). Of course, our ability to shuffle cards depends on practice and agility. The model produces shuffles with single cards being dropped about $1/2$ of the time, pairs of cards being dropped about $1/4$ of the time, and i cards blocks being dropped about $1/2^i$ of the time. Professional dealers drop single cards 80% of the time, pairs about 18% of the time and hardly ever drop 3 or more cards. Less sophisticated card handlers drop single cards about 60% of the time. Further discussion is in Diaconis (1982) or Epstein (1977).

It is not clear if neater shuffling makes for a better randomization mechanism. After all, eight perfect shuffles bring a deck back to order. Diaconis, Kantor, and Graham (1983) contains an extensive discussion of the mathematics of perfect shuffles, giving history and applications to gambling, computer science and group theory.

The shuffle analyzed above is the most random among all single shuffles with a given distribution of cut size, being uniform among the possible outcomes. It may therefore serve as a lower bound; any less uniform shuffle might take at least as long to randomize things. Further discussion is in Mellish (1973).

REMARK (f). One may ask, "Does it matter?" It seems to many people that if a deck of cards is shuffled 3 or 4 times, it will be quite mixed up for practical purposes with none of the esoteric patterns involved in the above analysis coming in.

Magicians and card cheats have long taken advantage of such patterns. Suppose a deck of 52 cards in known order is shuffled 3 times and cut arbitrarily in between these shuffles. Then a card is taken out, noted and replaced in a different position. The noted card can be determined with near certainty! Gardner (1977) describes card tricks based on the inefficiency of too few riffle shuffles.

Berger (1973) describes a different appearance of pattern. He compared the distribution of hands at tournament bridge before and after computers were used to randomize the order of the deck. The earlier, hand shuffled, distribution showed noticeable patterns (the suit distributions were too near "even" 4333) that a knowledgeable expert could use.

It is worth noting that it is not totally trivial to shuffle cards on a computer. The usual method, described in Knuth (1981), goes as follows. Imagine the n cards in a row. At stage i, pick a random position between i and n and switch the card at the chosen position with the card at position i. Carried out for $1 \leqslant i \leqslant n - 1$, this results in a uniform permutation. In the early days of computer randomization, we are told that Bridge Clubs randomized by choosing about 60 random transpositions (as opposed to 51 carefully randomized transpositions). As the analysis of

Diaconis and Shahshahani (1981) shows, 60 is not enough.

REMARK (g). While revising this paper we noted the following question and answer in a newspaper bridge column ("The Aces", by Bobby Wolff).

Q: How many times should a deck be shuffled before it is dealt? My fellow players insist on at least seven or eight shuffles. Isn't this overdoing it?

A: The laws stipulate that the deck must be "thoroughly shuffled". While no specific number is stated, I would guess that five or six shuffles would be about right; seven or eight would not be out of order.

5. Other Techniques and Open Problems. A number of other natural random walks admit elegant analyses with strong uniform times. For example, Andre Broder (1985) has given stopping times for simple random walk on the "cube" Z_2^d, and for the problem of randomizing n cards by random transpositions. We can similarly analyze nearest neighbor random walks on a variety of 2 point homogeneous spaces. It is natural to inquire if a suitable stopping time can always be found. This problem is analyzed in Aldous and Diaconis (1985): let us merely state two results.

We need to introduce a second notion of distance to the uniform distribution. Let Q be a probability on a finite group G. The *separation of Q^{k*}* to the uniform distribution U after k steps is defined as

$$s(k) = |G| \max_g \{U(g) - Q^{k*}(g)\}.$$

Clearly $0 \leqslant s(k) \leqslant 1$ with $s(k) = 0$ if and only if $Q^{k*} = U$. The separation is an upper bound for the variation distance:

$$d(k) \leqslant s(k)$$

because

$$\|Q^{k*} - U\| = \sum_{g:\, Q^{k*}(g) < U(g)} \{U(g) - Q^{k*}(g)\}.$$

The following result generalizes Lemma 1.

THEOREM 6. *If T is a strong uniform time for the random walk generated by Q on G, then*

$$s(k) \leqslant P(T > k); \quad \text{all } k \geqslant 0. \tag{5.1}$$

Conversely, for every random walk there exists a randomized strong uniform time T such that (5.1) holds with equality.

While separation and variation distance can differ, for random walk problems there is a sense in which they only differ by a factor of 2. For $0 < \varepsilon < \frac{1}{4}$, define

$$\phi(\varepsilon) = 1 - (1 - 2\varepsilon^{1/2})(1 - \varepsilon^{1/2})^2$$

and observe that $\phi(\varepsilon)$ decreases as ε decreases, and $\phi(\varepsilon) \sim 4\varepsilon^{1/2}$ as $\varepsilon \to 0$.

THEOREM 7. *For any distribution Q on any finite group G,*

$$s(2k) \leqslant \phi(2d(k)): k \geqslant 1, \text{provided } d(k) < \frac{1}{8}.$$

Thus, if k steps suffice to make the variation distance small, at most $2k$ steps are required to make the separation small.

Coupling is a probabilistic technique closely related to strong uniform times which achieves the exact variation distance. The coupling technique applies to Markov chains far more general than random walks on groups: see Griffeath (1975, 1978), Pitman (1976), Athreya and Ney (1977).

Random walk involves repeated convolution and it is natural to try to use Fourier analysis or its non-commutative analog, group representation. Such techniques can sometimes give very sharp

bounds. Letac (1981) and Takács (1982) are readable surveys. Diaconis and Shahshahani (1981, 1984) present further examples. Robbins and Bolker (1981) use other techniques.

Despite this range of available techniques, there are some shuffling methods for which we do not have good results on how many shuffles are needed; for example:

(i) Riffle shuffles where there is a tendency for successive cards to be dropped from opposite hands.

(ii) *Overhand shuffle*. The deck is divided into K blocks in some random way, and the order of the blocks is reversed.

(iii) *Semi-random transposition*. At the kth shuffle, transpose the kth card (counting modulo n) with a uniform random card.

From a theoretical viewpoint, there are interesting questions concerning the cut-off phenomenon. This occurs in all the examples we can explicitly calculate, but we know no general result which says that the phenomenon must happen for all "reasonable" shuffling methods.

Acknowledgment. We thank Brad Efron, Leo Flatto, and Larry Shepp for help with Example 1, and Jim Reeds for help with Section 4.

References

D. Aldous, Markov chains with almost exponential hitting times, Stochastic Proc. Appl., 13 (1982) 305–310.

______, Random walks on finite groups and rapidly mixing Markov chains, Séminaire de Probabilités, XVII (1983a) 243–297.

______, Minimization algorithms and random walk on the d-cube, Ann. Probab., 11 (1983b) 403–413.

D. J. Aldous and P. Diaconis, Uniform stopping times for random walks on groups, 1985. In preparation.

K. B. Athreya and P. Ney, A new approach to the limit theory of recurrent Markov chains, Trans. Amer. Math. Soc., 1977.

K. B. Athreya, D. McDonald, and P. Ney, Limit theorems for semi-Markov processes and renewal theory for Markov chains, Ann. Probab., 6 (1978) 788–797.

P. Berger, On the distribution of hand patterns in bridge: man-dealt versus computer-dealt, Canad. J. Statist., 1 (1973) 261–266.

E. Borel and A. Cheron, Théorie Mathématique du Bridge, 2nd ed., Gauthier Villars, Paris, 1955.

A. Broder, unpublished, Stanford University, 1985.

F. K. Chung, P. Diaconis, and R. L. Graham, Random walks arising from random number generation, Technical Report #212, Stanford University, 1983 (to appear, Ann. Probab.).

P. Diaconis (1982), On the use of group representations in probability and statistics, Typed Lecture Notes, Department of Statistics, Harvard University. To appear, Institute of Mathematical Statistics.

P. Diaconis, W. Kantor, and R. L. Graham, The mathematics of perfect shuffles, Advances in Applied Math., 4 (1983) 175–196.

P. Diaconis and M. Shahshahani, Generating a random permutation with random transpositions, Z. Wahrsch. Verw. Gebiete, 57 (1981) 159–179.

______, Factoring probabilities on compact groups, Technical Report #178, Stanford University (also TR#PH-9, Harvard Univ.) 1986 (to appear, Proc. Amer. Math. Soc.).

______, Products of random matrices and random walks on groups. To appear, Proc. Conference on Random Matrices (J. Cohen, H. Kesten, C. Newman, eds.), Amer. Math. Soc., Providence, RI, 1984.

J. R. Donner and V. R. R. Uppulini, A Markov chain structure for riffle shuffling, SIAM J. Appl. Math., 18 (1970) 191–209.

R. Epstein, The Theory of Gambling and Statistical Logic, Revised ed., Academic Press, New York, 1977.

W. Feller, An Introduction to Probability and Its Applications, Vol. I, 3rd ed., Wiley, New York, 1968.

L. Flatto, A. Odlyzko, and D. Wales, Random shuffles and group representations, Ann. Probab., 13 (1985) 181–193.

M. Gardner, Mathematical Magic Show, Knopf, New York, 1977.

E. W. Gilbert, Theory of Shuffling, Bell Laboratories Technical Memorandum, Murray Hill, NJ, 1955.

U. Grenander, Probability on Algebraic Structures, Wiley, New York, 1963.

D. Griffeath, A maximal coupling for Markov chains, Z. Wahrsch. Verw. Gebiete, 31 (1975) 95–106.

______, Coupling methods for Markov processes, in Studies in Probability and Ergodic Theory, Adv. in Math., Supplementary Studies Vol. 2 (J. C. Rota ed.), (1978) 1–43.

H. Heyer, Probability Measures on Locally Compact Groups, Springer, Berlin, 1977.

M. Iosifescu, Finite Markov Chains and Their Applications, Wiley, New York, 1980.

M. Kac, Random walk and the theory of Brownian motion, this MONTHLY, 54 (1947) 369–391.

B. M. Kloss, Probability distributions on bicompact topological groups, Theory Probab. Appl., 4 (1959) 237–270.

D. Knuth, The Art of Computer Programming, Vol. 1, 2nd ed., Addison-Wesley, Reading, MA, 1973.

______, The Art of Computer Programming, Vol. 2, 2nd ed., Addison-Wesley, Reading, MA, 1981.

D. D. Kosambi and U. V. R. Rao, The efficiency of randomization by card shuffling, J. Roy. Statist. Soc. Ser. A, 128 (1958) 223–233.

G. Letac, Problèmes Classiques de Probabilité sur un couple de Gelfand, Springer Lecture Notes 861, Springer, New York, 1981.

A. A. Markov, Extension of the law of large numbers to dependent events (Russian), Bull. Soc. Phys. Math. Kazan (2) 15 (1906) 135–156.

P. Matthews, Covering problems and random walks on groups, Ph.D. dissertation, Department of Statistics, Stanford University, 1985.

M. J. Mellish, Optimal Card-Shuffling, Eureka, No. 36 (1973) 9–10.

J. Pitman, On coupling of Markov chains, Z. Wahrsch. Verw. Gebiete, 35 (1976) 315–322.

H. Poincaré, Calcul Probabilités, Gauthier-Villars, Paris, 1912.

J. Reeds, Unpublished manuscript, 1981.

D. P. Robbins and E. D. Bolker, The bias of three pseudo random shuffles, Aequationes Math., 22 (1981) 268–292.

L. Takács, Random walks on groups, Linear Algebra Appl., 43 (1982) 49–67.

E. Thorp, Nonrandom shuffling with applications to the game of Faro, J. Amer. Statist. Assoc., 68 (1973) 842–847.

WHAT IS A DIFFERENTIAL? A NEW ANSWER FROM THE GENERALIZED RIEMANN INTEGRAL

SOLOMON LEADER

Mathematics Department, Hill Center, Busch Campus, Rutgers University, New Brunswick, NJ 08903

Unlike derivatives which gained a solid basis in Cauchy's theory of limits, differentials found no effective accommodation with the rising level of rigor in calculus. Justly castigated by Berkeley as "ghosts of departed quantities", differentials clung fortuitously to the notational niche in calculus created for them by Leibniz. In this century they came to be presented as functionals on tangent spaces, a constricted role that made them respectable but evaded the issue of their wider role in integration. The resurrection of infinitesimals by nonstandard analysis rekindled interest in Leibniz' original concept of differential.

We present here a completely new approach to differentials in one dimension. This approach is motivated by the following considerations: (i) differentials spring directly from the integration process, (ii) the utility of differentials in integration extends beyond conventional differential forms, (iii) a viable theory of differentials is readily attainable by standard analysis, and (iv) the generalized Riemann integral fills a vital gap in analysis and should have an innovative impact on our calculus and real variables courses. In the theory expounded here differentials on a 1-cell K form a Riesz space (lattice-ordered linear space). So for each differential σ we have the differentials

$$|\sigma| = \sigma \vee -\sigma, \quad \sigma^{+} = \sigma \vee 0, \quad \text{and} \quad \sigma^{-} = (-\sigma)^{+} = -(\sigma \wedge 0)$$

Solomon Leader: I wrote my Ph.D. thesis in analysis at Princeton in 1952 under the late Salomon Bochner. For the past 33 years I have been at Rutgers figuring out how calculus should be taught. My main interests have been in measure theory, integration, proximity spaces, and fixed points. In warm weather my favorite diversion is body-surfing off Long Beach Island. My wife and I enjoy snorkeling in the Virgin Islands and welcome any excuse to visit Switzerland.

with

$$\sigma = \sigma^+ - \sigma^- \quad \text{and} \quad |\sigma| = \sigma^+ + \sigma^-.$$

Every function x on K has an integrable differential dx. Moreover, if σ is a differential, then so is $x\sigma$. Thus, formulas such as

$$dz = ydx, \quad dz = y|dx|, \quad dy = (dx)^+, \quad |ydx| \leqslant dz, \quad \text{and} \quad d(xy) = xdy + ydx$$

have well-defined meanings. We can also apply Lipschitz functions to differentials to get differentials such as

$$\sigma = (dx^2 + dy^2)^{1/2}.$$

Our theory of differentials is based on a modification of Kurzweil's generalized Riemann integral [4]. We shall give its definition and summarize its relevant properties in conjunction with our development of differentials in one dimension. A detailed exposition [7] complete with proofs and technical details will appear elsewhere. Successful extension to higher dimensions seems to rest on the right choice of regularity restrictions on the eccentricity of the cells in partitions used to define the integral. (See [10], [5], [6].) An excellent elementary introduction to the generalized Riemann integral is [8]. An extensive bibliography can be found in [11]. For basic facts about Riesz spaces (vector lattices) see [1].

1. Basic Definitions. A *function* is any mapping with values in $\boldsymbol{R} = (-\infty, \infty)$. A *cell* is a 1-cell (closed bounded nondegenerate interval in $\boldsymbol{R}$). A *figure* is a nonvoid union of finitely many cells. Two figures *overlap* if their intersection contains a cell. A finite set of cells *partitions* its union if no two cells overlap. A *tagged cell* (I, t) is a cell I with a selected end-point t. (Here we depart from Kurzweil [4] who lets t be any point in I. McShane [9] puts no restriction on the tag so he gets only absolutely convergent integrals.) A *division* $\mathscr{F}$ of a figure F is a finite set of tagged cells which partitions F. A *gauge* on F is a function α on F with $\alpha(t) > 0$ for all t. (Here the generalized Riemann integral crucially differs from the Riemann integral which in effect uses only constant gauges.) A tagged cell (I, t) is *α-fine* if the length of I is less than $\alpha(t)$. An *α-division* is a division whose members are α-fine. (For any gauge α on a cell K the existence of an α-division of K can be shown by a Heine-Borel argument.) A *summant* S over a figure F is a function $S(I, t)$ on the set of all tagged cells contained in F. S is a *cell summant* if $S(I, p) = S(I, q)$ for every cell $I = [p, q]$ in F. For such S we can denote $S(I, t)$ by $S(I)$ since the tag is irrelevant. For y a function on F and S a summant over F, let yS be the summant with value $y(t)S(I, t)$ at (I, t). The most common summants are of this form with S a cell summant Δx defined for a function x on F by

$$\Delta x(I) = x(q) - x(p)$$

for $I = [p, q]$. We shall also deal with $y|\Delta x|$ which assumes the value $y(t)|\Delta x(I)|$ at (I, t). For S any summant over F each division $\mathscr{F}$ of F yields a *Riemann sum* $\Sigma(S, \mathscr{F})$, the sum of $S(I, t)$ over all (I, t) in $\mathscr{F}$. Define the *lower* and *upper integrals* with values in $[-\infty, \infty]$ as follows. For each gauge α on F define $\underline{\Sigma}(S, \alpha)$ to be the infimum, $\overline{\Sigma}(S, \alpha)$ the supremum, of all sums $\Sigma(S, \mathscr{F})$ with $\mathscr{F}$ any α-division of F. Then the lower integral is

$$\underline{\int} S = \sup_\alpha \underline{\Sigma}(S, \alpha)$$

and the upper integral is

$$\overline{\int} S = \inf_\alpha \overline{\Sigma}(S, \alpha),$$

where α runs through all the gauges on F. Whenever the lower and upper integrals are equal define the *integral* $\int S = \underline{\int} S = \overline{\int} S$ in $[-\infty, \infty]$. S is *integrable* whenever $\int S$ exists and is finite. It

can be shown that S is integrable with $\int S = c$ if and only if given $\varepsilon > 0$, there exists a gauge α on F such that

$$|c - \Sigma(S, \mathscr{F})| < \varepsilon$$

for every α-division $\mathscr{F}$ of F. A necessary and sufficient condition for S to be integrable is the Cauchy criterion: given $\varepsilon > 0$ there exists a gauge α on F such that

$$|\Sigma(S, \mathscr{F}_1) - \Sigma(S, \mathscr{F}_2)| < \varepsilon$$

for all α-divisions $\mathscr{F}_1, \mathscr{F}_2$ of F. The integrable summants over F form an ordered linear function space on which the integral acts as a positive linear functional. But this function space is not a Riesz space. It fails to be a lattice since S may be integrable but not $|S|$. S is *absolutely integrable* if both S and $|S|$ are integrable.

2. Differentials. The function space $\mathbf{S}$ of all summants S over F is a Riesz space under the usual operations on functions. The set $\mathbf{N}$ of all absolutely integrable S in $\mathbf{S}$ with $\int|S| = 0$ is a *Riesz ideal*. That is, $\mathbf{N}$ is a linear subspace which is *solid* in $\mathbf{S}$: If $S \in \mathbf{S}$, $T \in \mathbf{N}$, and $|S| \leqslant |T|$, then $S \in \mathbf{N}$. Thus $\mathbf{D} = \mathbf{S}/\mathbf{N}$ is a Riesz space with the linear and lattice operations transferred homomorphically from $\mathbf{S}$ to $\mathbf{D}$. A *differential* σ on F is any element of $\mathbf{D}$. Specifically, σ is an equivalence class $[S]$ of summants over F under the equivalence $S \sim T$ defined by $\int|S - T| = 0$. For $\rho = [R]$ and $\sigma = [S]$ the homomorphism gives $\rho + \sigma = [R + S]$, $c\sigma = [cS]$ for any constant c, and $|\sigma| = [|S|]$. Hence

$$\rho \wedge \sigma = [R \wedge S], \quad \rho \vee \sigma = [R \vee S], \quad \sigma^+ = [S^+], \quad \text{and} \quad \sigma^- = [S^-].$$

We can transfer the differential ordering $\rho \leqslant \sigma$ defined by $(\rho - \sigma)^+ = 0$ to representative summants, defining $R \lesssim S$ to be $\int(R - S)^+ = 0$. Then $R \sim S$ if, and only if, both $R \lesssim S$ and $S \lesssim R$. $R \lesssim S$ implies $\underline{\int} R \leqslant \underline{\int} S$ and $\overline{\int} R \leqslant \overline{\int} S$. So $R \sim S$ implies $\underline{\int} R = \underline{\int} S$ and $\overline{\int} R = \overline{\int} S$. Thus we can effectively define the lower and upper integrals of any differential $\sigma = [S]$ on F by $\underline{\int}\sigma = \underline{\int} S$ and $\overline{\int}\sigma = \overline{\int} S$. Define $\int\sigma = \underline{\int}\sigma = \overline{\int}\sigma$ whenever the latter two are equal. Call σ *integrable* (*absolutely integrable*) whenever S is integrable (absolutely integrable).

Every Lipschitz function f on $\mathbf{R}^n$ induces a mapping $\tilde{f}$ on $\mathbf{D}^n$ into $\mathbf{D}$ defined by

$$\tilde{f}(\sigma_1, \dots, \sigma_n) = [f(S_1, \dots, S_n)]$$

for $\sigma_i = [S_i]$ with $i = 1, \dots, n$. Since all norms on $\mathbf{R}^n$ are equivalent, the Lipschitz condition yields a constant $c > 0$ such that for all $\mathbf{p}, \mathbf{q}$ in $\mathbf{R}^n$

$$|f(\mathbf{p}) - f(\mathbf{q})| \leqslant c\|\mathbf{p} - \mathbf{q}\|_1,$$

where

$$\|\mathbf{r}\|_1 = |r_1| + \cdots + |r_n| \quad \text{for} \quad \mathbf{r} = (r_1, \dots, r_n).$$

This makes the definition of $\tilde{f}$ effective: $f(S_1', \dots, S_n') \sim f(S_1, \dots, S_n)$ if $S_i' \sim S_i$ for all i. In particular the definition applies to any norm f on $\boldsymbol{R}^n$. Thus for example, if $\sigma_1, \dots, \sigma_n$ are differentials on F, then so is $(|\sigma_1|^p + \cdots + |\sigma_n|^p)^{1/p}$ for $p \geqslant 1$.

For any function x on F define the differential $dx = [\Delta x]$. This is not to be confused with that bogus formula "$dx = \Delta x$" used to palm off the approximation formula $\Delta y \doteq y' \Delta x$ as the differential relation $dy = y'dx$. As we shall see in (5), for any function x the differential dx is integrable.

To show the utility of our definition of differential we must examine the integral more closely.

3. Basic Properties of the Integral. Hereafter we may denote $\int$ by $\int_F$ wherever the underlying figure F may be ambiguous. We also revert to the classical notation $\int_a^b$ for $\int_{[a,b]}$ with the convention $\int_a^a = 0$. A point c in f is a tag in every α-division of F for α sufficiently small,

specifically if $\alpha(t) < |t - c|$ for all $t \neq c$. Thus, since tags are endpoints of tagged cells, we have the additivity theorem (1).

(1) *Let σ be a differential on the union C of two nonoverlapping figures A, B. Then*

$$\underline{\int}_C \sigma = \underline{\int}_A \sigma + \underline{\int}_B \sigma \quad \text{and} \quad \overline{\int}_C \sigma = \overline{\int}_A \sigma + \overline{\int}_B \sigma$$

ignoring the indeterminate form $\infty - \infty$. Hence if σ is integrable on both A and B, then σ is integrable on C and

$$\int_C \sigma = \int_A \sigma + \int_B \sigma.$$

Application of the Cauchy criterion for gauge limits gives uniform integrability on subfigures. Specifically we have (2).

(2) *Let σ be integrable on F, let $S \in \sigma$, and let α be a gauge on F such that $|\Sigma(S, \mathscr{F}) - \int_F \sigma| \leqslant \varepsilon$ for every α-division $\mathscr{F}$ of F. Then σ is integrable on every figure E contained in F and $|\Sigma(S, \mathscr{E}) - \int_E \sigma| \leqslant \varepsilon$ for every α-division $\mathscr{E}$ of E.*

With a bit of ingenuity (1), (2) give (3).

(3) *Given σ integrable on F define $S_*(I) = \int_I \sigma$ for every cell I contained in F. Then the cell summant $S_* \in \sigma$.*

So $S_*{}^+ \in \sigma^+$ which immediately gives (4).

(4) *For σ integrable on F $\sigma \geqslant 0$ if and only if $\int_I \sigma \geqslant 0$ for every cell I contained in F. So $\sigma = 0$ if and only if $\int_I \sigma = 0$ for every cell I contained in F.*

Since $\Sigma(\Delta x, \mathscr{I}) = \Delta x(I)$ for every partition (or division) $\mathscr{I}$ of a cell I, we have (5).

(5) *Let x be a function on F. Then $\Delta x(I) = \int_I dx$ for every cell I contained in F.*

Given σ integrable on F define x on each component $[a, b]$ of F by $x(t) = \int_a^t \sigma$. Then $\Delta x = S_*$ by (1), (5). So $dx = \sigma$ by (3). By (5) this gives (6).

(6) *σ is integrable on F if and only if $\sigma = dx$ for some function x on F.*

Since gauge-limit divisions with endpoint tags ultimately refine any given partition, we get a nice formulation in terms of differentials for total variation.

(7) *For every function x on $[a, b]$ the integral $\int_a^b |dx|$ exists in $[0, \infty]$ and equals the total variation of x. For x of bounded variation define $v(t) = \int_a^t |dx|$, $y = \frac{1}{2}(v + x)$, and $z = \frac{1}{2}(v - x)$ to get the Jordan decomposition $x = y - z$ with $v = y + z$, $dv = |dx|$, $dy = (dx)^+$, and $dz = (dx)^-$.*

Thus if σ is integrable on F, then $\int|\sigma|$ exists in $[0; \infty]$ by (5), (6), and (7). (3) is a differential formulation of Henstock's Lemma. Henstock and Kurzweil independently developed the generalized Riemann integral. Henstock [2] even used endpoint tags.

Our next objective is to show that differentials can be multiplied by functions.

4. Monotone Convergence. Our next result is deeper than any of the preceding results. Indeed, it leads eventually to the conclusion that the Lebesgue-Stieltjes integral is essentially the restriction of our integral to the case $y\,dx$ with both dx and $y\,dx$ absolutely integrable. We state this theorem without any discussion of its proof.

(8) *Let $S \geqslant 0$ be a summant over F. Let v and $v_1, \dots, v_n, \dots$ be nonnegative functions on F such that $v \leqslant \sum_{n=1}^{\infty} v_n$. Then $\overline{\int} vS \leqslant \sum_{n=1}^{\infty} \overline{\int} v_n S$. If moreover $\int v_n S$ exists for all n and $v = \sum_{n=1}^{\infty} v_n$, then $\int vS = \sum_{n=1}^{\infty} \int v_n S$.*

A consequence of (8) that we need is (9).

(9) *Let $T \sim 0$ over F and let y be a function on F. Then $yT \sim 0$.*

To get (9) from (8) let $S = |T|$, $v = |y|$, and $v_n = 1$. Then

$$v < \infty = \sum_{n=1}^{\infty} v_n.$$

So (8) gives

$$\overline{\int} |yT| \leqslant \sum_{n=1}^{\infty} \int |T| = 0.$$

We can now define $y\sigma = [yS]$ for $\sigma = [S]$ on F and y any function on F. Given $S' \sim S$ apply (9) with $T = S' - S$ to conclude that $yS' \sim yS$. So the definition of $y\sigma$ is effective. It coincides with scalar multiplication when y is constant. Also $(yz)\sigma = y(z\sigma)$ and $|y\sigma| = |y||\sigma|$.

For A a subset of $\mathbf{R}$ let 1_A be the *indicator* of A. That is, 1_A is the function on $\mathbf{R}$ with $1_A(t) = 1$ for all t in A, and $1_A(t) = 0$ elsewhere. For σ a differential on F and A a subset of F call A *σ-null* if $1_A \sigma = 0$. A condition on points of F holds *a.e.* σ if it holds on the F-complement of some σ-null subset A of F. We can now formulate a monotone convergence theorem whose proof is based on (8).

(10) *Let $\sigma \geqslant 0$ be a differential on F. Let $0 \leqslant y_1 \leqslant \cdots \leqslant y_n \leqslant \cdots$ be a sequence of functions on F such that each $y_n \sigma$ is integrable and $\int y_n \sigma \nearrow q < \infty$. Then the set of all t where $y_n(t) \nearrow \infty$ is σ-null. If y is any function of F such that $y_n \nearrow y$ a.e. σ, then $\int y\sigma = q$.*

A corollary to (10) is (11).

(11) *Let σ be a differential on F and let y be a function on F. Then $y\sigma = 0$ if and only if $y = 0$ a.e. σ.*

The foregoing properties suffice to show that our integral subsumes the Lebesgue-Stieltjes integral on a cell $K = [a, b]$. They render superfluous the extension techniques heretofore required to construct a Borel measure on K from its distribution function. Indeed, given any isotone ($dx \geqslant 0$) function x on K, properties (1), (2), (8), (10) show that the subsets E of K with $1_E dx$ integrable form a sigma-algebra $\mathbf{A}$ on which $m(E) = \int_K 1_E dx$ defines a measure, m on K. Clearly,

$$\int_K 1_I dx = x(q+) - x(p-)$$

for $I = [p, q]$ contained in K, with the convention

$$x(a-) = x(a), \quad x(b+) = x(b).$$

So every cell, hence every Borel set, in K belongs to $\mathbf{A}$. Thus m is a finite Borel measure on K.

5. Dampable Differentials. Since every summant S over F represents a differential σ, we have bagged some strange differentials in $\mathbf{D}$. We can exclude pathological types by introducing some restrictions. A function w on F is a *damper* for σ if $w(t) > 0$ for all t and $w\sigma$ is absolutely integrable. If such w exist, we call σ *dampable*. This property is roughly analogous to sigma-finiteness in measure theory. The class of functions x on F with dx dampable is quite large. It includes all functions of bounded variation (with damper 1) and, as we shall conclude from (16), all differentiable functions. But there do exist continuous functions x with dx not dampable. An open question is: Must dx be dampable if $|dx|$ is?

6. Tag-Null Differentials. Although the discrete aspect of differentials is of interest (see [7]), the classical calculus is concerned only with the continuous aspect (no point charges). A differential σ on F is *tag-null* if every singleton, hence by (8) every countable set, in F is σ-null. Tag-null differentials are truly "ghosts of departed quantities" in that $\sigma = [S]$ is tag-null if and only if $S(I, t) \to 0$ as $I \to t$ in F for all t. The convergence $I \to t$ in F pertains to all tagged cells (I, t) in F with tag t as the length of I converges to 0. For the case of integrable differentials, dx is tag-null if and only if x is continuous.

7. The Chain Rule for Differentials.. We can now show that well-behaved differentials comply with the basic rules of calculus. Our next result shows that derivatives are always integrable, whether or not they are Lebesgue-summable. Specifically, if x is a continuous function on a cell K with $|dx|$ dampable and $z = f(x)$ with f differentiable, then $dz = f'(x)dx$. This is the case $n = 1$ of (12).

(12) *For $i = 1, \dots, n$ let x_i be a continuous function on $K = [a, b]$ with $|dx_i|$ dampable. Let $\mathbf{C}$ be the curve in $\mathbf{R}^n$ defined by $\mathbf{x}(t) = (x_1(t), \dots, x_n(t))$ for t in K. Let f be a function on a neighborhood of $\mathbf{C}$ in $\mathbf{R}^n$ such that f is differentiable at every point $\mathbf{x}$ in $\mathbf{C}$. Let $z(t) = f(\mathbf{x}(t))$ and $y_i(t) = \dfrac{\partial f}{\partial x_i}(\mathbf{x}(t))$. Then $dz = \sum_{i=1}^n y_i dx_i$.*

Since this is a typical application of dampability and tag-nullity, let us give a proof of (12). For f differentiable the gradient $\mathbf{y} = \left(\dfrac{\partial f}{\partial x_1}, \dots, \dfrac{\partial f}{\partial x_n}\right)$ exists at each point $\mathbf{x}$ in $\mathbf{C}$ and

$$(*) \qquad f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - \mathbf{y} \cdot \mathbf{h} = o(\|\mathbf{h}\|_1) \quad \text{as} \quad \mathbf{h} \to \mathbf{0} \quad \text{in} \quad \mathbf{R}^n.$$

Given a tagged cell (I, t) in K whose other endpoint is $t + q$, let $\mathbf{x} = \mathbf{x}(t)$ and

$$\mathbf{h} = \mathbf{x}(t + q) - \mathbf{x}(t) = (\operatorname{sgn} q)\Delta\mathbf{x}(I),$$

where $\Delta\mathbf{x} = (\Delta x_1, \dots, \Delta x_n)$. By continuity of $\mathbf{x}(t)$, $\mathbf{h} \to \mathbf{0}$ as $q \to 0$; that is, as $I \to t$. Note that

$$f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) = (\operatorname{sgn} q)\Delta z(I).$$

Let w_i be a damper for $|dx_i|$. Given $\varepsilon > 0$, choose a function w on K with $0 < w(t) < \varepsilon w_i(t)$ for all t in K and $i = 1, \ldots, n$. Then choose a gauge α on K small enough so that $(*)$ gives

$$|\Delta z(I) - \sum_{i=1}^{n} y_i(t)\Delta x_i(I)| \leqslant w(t)\|\Delta \mathbf{x}(I)\|_1$$

for all α-fine (I, t) in K. This implies

$$|dz - \sum_{i=1}^{n} y_i dx_i| \leqslant w\sum_{i=1}^{n}|dx_i| \leqslant \varepsilon\sum_{i=1}^{n} w_i|dx_i|.$$

Since the last sum is integrable and ε is arbitrary, $dz - \sum_{i=1}^{n} y_i dx_i = 0$.

From (12) we can get the classical formulas in the calculus of differentials. For example let u, v be continuous functions on K with $|du|, |dv|$ dampable. For $f(u, v) = uv$ (12) gives the product rule $d(uv) = udv + vdu$ used for integration by parts. If v has no zeros in K, we can apply (12) with $f(u, v) = u/v$ to get the quotient rule $d(u/v) = (vdu - udv)/v^2$.

For the special case $n = 1$ in (12) the conclusion $dz = ydx$ holds under less stringent conditions than those demanded by (12). We do not need a functional relation f between z and x to define the derivative dz/dx. Moreover, this derivative need not exist at every point in K to give $dz = ydx$. To delve into this we need the concept of absolute continuity for differentials.

8. Absolute Continuity. Let ρ, σ be differentials on F. ρ is *absolutely continuous* with respect to σ if every σ-null set is ρ-null. This is quite weak in the absence of other conditions. So we define ρ to be *strongly absolute continuous* with respect to σ if ρ, σ are absolutely integrable and given $\varepsilon > 0$ there exists $\delta > 0$ such that $\int_G|\rho| < \varepsilon$ for every figure G contained in F with $\int_G|\sigma| < \delta$. For tag-null, absolutely integrable ρ, σ the two forms of absolute continuity are equivalent. Therefore if ρ and σ are tag-null and dampable with dampers v, w, respectively, ρ is absolutely continuous with respect to σ if and only if $v\rho$ is strongly absolutely continuous with respect to $w\sigma$.

9. The Fundamental Theorem of Calculus for $dz = ydx$. For x, y, z functions on a cell $K = [a, b]$, the differential relation $dz = ydx$ says concisely by (5) that ydx is integrable and $z(t) = z(a) + \int_a^t ydx$ for all t in K. The fundamental theorem of calculus relates this to $dz/dx = y$, where the derivative is defined by

$$\frac{dz}{dx}(t) = \lim_{h\to 0}\frac{z(t+h) - z(t)}{x(t+h) - x(t)}$$

with t and $t + h$ in K. Existence of this limit requires that $x(t + h) \neq x(t)$ ultimately as $h \to 0$. This generalized derivative is the natural counterpart to our differentials. It plays an anonymous role in Cauchy's Generalized Theorem of the Mean and satisfies the chain rule for derivatives. dz/dx gives the slope of the curve in the (x, z)-plane traced by a particle with position $(x(t), z(t))$ at time t. Its existence does not require the existence of the velocity $(\dot{x}, \dot{z})$. For example, with $x = |t|$ and $z = t^3$ on $[-1,1]$ we have $dz/dx = 3t|t|$ even at $t = 0$ where $\dot{x}$ does not exist. Of course, wherever the velocity does exist and $\dot{x} \neq 0$ we have $dz/dx = \dot{z}/\dot{x}$ by the chain rule. (Since we are using the traditional Leibniz notation, we must emphasize that dz/dx is not a quotient of differentials from our point of view.) Note that in terms of summants

$$\frac{dz}{dx}(t) = \lim_{I\to t}\frac{\Delta z(I)}{\Delta x(I)},$$

where the cells I are contained in K and have t as an endpoint.

We can now formulate a nice version (13) of the fundamental theorem. The proof in [7] that (i) implies (ii) uses the Vitali Covering Theorem applied to the measure $w|dx|$ with w a damper for $|dx|$. That (ii) implies (iii) is trivial. The proof that (iii) implies (i) is similar to the proof of (12).

(13) *Let x, y, z, be functions on a cell K such that x is continuous and $|dx|$ dampable. Then the following are equivalent*:

(i) $dz = ydx$,

(ii) $\dfrac{dz}{dx} = y$ *a.e.* dx *and* dz *is absolutely continuous with respect to* dx,

(iii) $\dfrac{dz}{dx} = y$ *a.e.* $(dz - ydx)$.

A corollary to (13) is

(14) *Let x, z be continuous functions on K with $|dx|$ dampable. Let $\dfrac{dz}{dx}$ exist and be finite a.e. dz. Define $y(t) = \dfrac{dz}{dx}(t)$ wherever the derivative exists and $y(t) = 0$ elsewhere. Then $dz = ydx$.*

To get (14) from (13) let A be the set of all t where the relation $\dfrac{dz}{dx}(t) = y(t)$ fails, that is, where $\dfrac{dz}{dx}(t)$ does not exist. Then $1_A dz = 0$ by hypothesis, and $1_A y = 0$ by the definition of y. So $1_A(dz - ydx) = 0$ which gives (iii) in (13), hence (i).

With suitable restrictions on both x and z these results can be combined with the Radon–Nikodym Theorem to get

(15) *Let x, z be continuous functions on K with dx, dz dampable. Then the following are equivalent*:

(i) $dz = ydx$ *for some function y on K*,

(ii) dz *is absolutely continuous with respect to* dx,

(iii) $\dfrac{dz}{dx}$ *exists and is finite* a.e. dz.

The results (13), (14), (15) also hold for $dz = y|dx|$ with dx replaced by $|dx|$ throughout, and

$$\frac{dz}{|dx|}(t) = \lim_{h\to 0} \frac{z(t+h) - z(t)}{|x(t+h) - x(t)|}.$$

The proof of our final result (16) uses the theory of Borel measureable functions (see [7]).

(16) *Let x, y, z be functions on K such that x is continuous, $|dx|$ is dampable, and $dz = ydx$. Then there exists a Borel measurable function y_0 on K such that $y_0 = y$* a.e. dx. *Also, z is continuous and $|dz|$ is dampable. If dx is dampable, then so is dz.*

Note that by (11) we get $y_0 dx = ydx = dz$ from (16). We can now see why dz is dampable for z any differentiable function on K. The identity $x(t) = t$ is continuous, and (14) gives $dz = z'dt$. Since $dt \geqslant 0$, dt is dampable (with damper 1). So dz is dampable by (16).

10. Conclusion. These selected results show the power, clarity, and economy inherent in differential formulations of integration properties. They testify to the vitality of the generalized Riemann integral whose potent messages give substance to our differential shorthand. A generation has passed since the emergence of this integral during which it has been largely ignored or disdained (see [3]). But the generalized Riemann integral truly fills a basic gap in integration theory.

References

1. D. H. Fremlin, Topological Riesz Spaces and Measure Theory, Cambridge University Press, Cambridge, 1974.

2. R. Henstock, Theory of Integration, Butterworths, London, 1963.

3. ______, The Lebesgue syndrome, Proc. 6th Summer Symposium, Real Anal. Exchange, 9 (1983–84) 96–110.

4. K. Jacobs, Measure and Integral (Appendix by J. Kurzweil), Academic Press, New York, 1978.

5. J. Jarnik and J. Kurzweil, A nonabsolutely convergent integral which admits C^1-transformations, Časopis Pěst. Mat., 109 (1984) 157–167.

6. J. Jarnik, J. Kurzweil, and Š. Schwabik, On Mawhin's approach to multiple nonabsolutely convergent integral, Časopis Pěst. Mat., 108 (1983) 356–380.

7. S. Leader, The concept of differential based on generalized Riemann integration (in preparation).

8. R. M. McLeod, The Generalized Riemann Integral, Carus Mathematical Monograph no. 20, Mathematical Association of America, Washington, DC, 1980.

9. E. J. McShane, Unified Integration, Academic Press, New York, 1983.

10. J. Mawhin, Generalized multiple Perron integrals and the Green-Goursat theorem for differentiable vector fields, Czechoslovak Math. J., 31 (1981) 614–632.

11. B. S. Thomson, Derivation bases on the real line I, II, Real Anal. Exchange, 8 (1982–83) 67–207, 278–442.

TRANSFORMATIONS LEAVING THE DETERMINANT OF CIRCULANT MATRICES INVARIANT

ALAN C. WILDE
University of Michigan, Ann Arbor, MI 48109

1. Circulant matrices. These are matrices such as

$$\begin{bmatrix} 2 & 5 & 8 & 9 \\ 9 & 2 & 5 & 8 \\ 8 & 9 & 2 & 5 \\ 5 & 8 & 9 & 2 \end{bmatrix}$$

in which the rows are obtained each from its predecessor by the application of a cyclic permutation. These matrices have been studied since the time of Catalan (1846). (See Muir [2].) They have an amazing variety of special properties. In particular, they are related to discrete Fourier transforms.

Thus the eigenvalues $\lambda_0, \lambda_1, \ldots, \lambda_{n-1}$ of the $n \times n$ complex circulant matrix X with first row $(x_0, x_1, \ldots, x_{n-1})$ have the discrete Fourier expansion (by Davis [1], p. 73)

$$\lambda_h = \sum_{j=0}^{n-1} \zeta^{hj} x_j \quad \text{for } 0 \leqslant h \leqslant n-1, \quad \text{where } \zeta = e^{2\pi i/n}, \tag{1}$$

and the corresponding eigenvectors are

$$\vec{b}_h = \mathrm{Col}(1, \zeta^h, \zeta^{2h}, \ldots, \zeta^{(n-1)h}) \text{ for } 0 \leqslant h \leqslant n-1, \tag{2}$$

where "$\mathrm{Col}(x_0, x_1, \ldots, x_{n-1})$" denotes the column vector with entries $x_0, x_1, \ldots, x_{n-1}$. The matrix X is diagonalized by the equation (Davis [1], p. 78)

$$B \,\mathrm{diag}(\lambda_0, \lambda_1, \ldots, \lambda_{n-1})\, B^{-1} = X, \tag{3}$$

A son of a physiologist, Alan Wilde received his B.S. and M.A. in mathematics from the University of Michigan. He works as an Independent Study Teacher for the University of Michigan Extension Service. His six published papers cover a diverse but interrelated range of mathematical topics, including geometry, algebra, logic, differential equations, and functional equations.

where "diag$(\lambda_0, \dots, \lambda_{n-1})$" denotes the diagonal matrix with entries $\lambda_0, \lambda_1, \dots, \lambda_{n-1}$ on the main diagonal, and B is the Vandermonde matrix

$$B = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \zeta & \zeta^2 & \dots & \zeta^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \zeta^{n-1} & \zeta^{2(n-1)} & \dots & \zeta^{(n-1)^2} \end{bmatrix}. \tag{4}$$

(The Fourier matrix equals $(1/\sqrt{n})B$.) The determinant of X is, as usual, the product of the eigenvalues and is hence equal to (Davis [1], p. 92)

$$\det X = \Delta = \prod_{h=0}^{n-1} \left(\sum_{j=0}^{n-1} \zeta^{hj} x_j \right). \tag{5}$$

The set of $n \times n$ complex circulant matrices X is a commutative algebra over $\mathbb{C}$. Let K be the circulant matrix with $x_0 = 0$, $x_1 = 1$, $x_2 = 0, \dots, x_{n-1} = 0$, i.e.,

$$K = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & 0 & \dots & 0 \end{bmatrix}. \tag{6}$$

Then $K^n = I$, and (Davis [1], p. 68)

$$X = \sum_{h=0}^{n-1} x_h K^h. \tag{7}$$

Let $E_{h,j}$ denote the complex $n \times n$ matrix with 1 in the $(h+1, j+1)$ spot and 0's elsewhere where $0 \leqslant h \leqslant n-1$ and $0 \leqslant j \leqslant n-1$. Then $\{E_{h,j} | 0 \leqslant h \leqslant n-1, 0 \leqslant j \leqslant n-1\}$ is a basis of all $n \times n$ complex matrices. Circulant matrices are of the form

$$X = \sum_{h=0}^{n-1} \sum_{j=0}^{n-1} x_j E_{h,h+j}, \quad h+j \text{ taken modulo } n. \tag{8}$$

From now on, we will let $\sum_{h,j}$ denote $\sum_{h=0}^{n-1} \sum_{j=0}^{n-1}$.

For reasons given later, let

$$X' = \sum_{h,j} x'_j E_{h,h+j}, \quad h+j \text{ taken modulo } n. \tag{9}$$

Then the eigenvalues of X' are

$$\lambda'_h = \sum_{j=0}^{n-1} \zeta^{hj} x'_j \quad \text{for} \quad 0 \leqslant h \leqslant n-1. \tag{10}$$

The determinant of X' is thus

$$\det X' = \Delta' = \prod_{h=0}^{n-1} \left(\sum_{j=0}^{n-1} \zeta^{hj} x'_j \right). \tag{11}$$

Let $\vec{X}$ and $\vec{X}'$ be the vectors

$$\vec{X} = \mathrm{Col}(x_0, x_1, \dots, x_{n-1}) \tag{12}$$

and

$$\vec{X}' = \mathrm{Col}(x_0', x_1', \dots, x_{n-1}'). \tag{13}$$

All these properties are used in my main theorem.

2. Main theorem. Wilfred Kaplan posed a problem: find all linear transformations on the top row-vector of X that leave the determinant of the circulant invariant. The solution is as follows.

THEOREM. *Let n be fixed. Let G be an $n \times n$ complex matrix of the form*

$$G = \frac{1}{n} \sum_{h,j} \left(\sum_{q=0}^{n-1} \zeta^{-h\sigma(q)+jq} \alpha_q \right) E_{h,j}, \tag{$*$}$$

where σ is a permutation on $\{0,1,\dots,n-1\}$ and $\alpha_0, \alpha_1, \dots, \alpha_{n-1} \in \mathbb{C}$ such that $\prod_{q=0}^{n-1} \alpha_q = 1$. Then for all $n \times n$ complex circulant matrices X, the substitution

$$\vec{X}' = G\vec{X} \tag{$**$}$$

determines an $n \times n$ circulant matrix X' such that

$$\det X = \det X'. \tag{$***$}$$

Conversely, if the substitution $(**)$ leads to $(***)$ for all $\vec{X}$, then G has the form $(*)$.

Proof of $(\Leftarrow)$. By equation (4),

$$B = \sum_{h,j} \zeta^{hj} E_{h,j}. \tag{14}$$

Then, as is well known,

$$B^{-1} = \frac{1}{n} \sum_{h,j} \zeta^{-hj} E_{h,j}. \tag{15}$$

Now let

$$\vec{\lambda} = \mathrm{Col}(\lambda_0, \lambda_1, \dots, \lambda_{n-1}), \tag{16}$$

and

$$\vec{\lambda}' = \mathrm{Col}(\lambda_0', \lambda_1', \dots, \lambda_{n-1}'). \tag{17}$$

Then by equations (1), (12), (13) and (14),

$$\vec{\lambda} = B\vec{X} \tag{18}$$

and

$$\vec{\lambda}' = B\vec{X}'. \tag{18$'$}$$

By equations (5), (11), (1), and (10),

$$\det X = \prod_{h=0}^{n-1} \lambda_h \tag{19}$$

and

$$\det X' = \prod_{h=0}^{n-1} \lambda_h'. \tag{20}$$

Thus condition $(***)$ is equivalent to the condition

$$\prod_{h=0}^{n-1} \lambda_h' = \prod_{h=0}^{n-1} \lambda_h. \tag{21}$$

Finding the set of G's such that $(***)$ holds corresponds to finding the corresponding transformations on the eigenvalues such that (21) holds. In detail, $(**)$ is $\vec{X}' = G\vec{X}$, and by (18) and (18′), this gives $\vec{\lambda}' = BGB^{-1}\vec{\lambda}$, or

$$(22) \qquad \lambda'_h = \sum_{j=0}^{n-1} \alpha_{hj}\lambda_j \quad \text{for} \quad 0 \leqslant h \leqslant n-1,$$

where $\alpha_{hj} \in \mathbf{C}$ for $0 \leqslant h \leqslant n-1$ and $0 \leqslant j \leqslant n-1$.

Suppose now that G is such that $(***)$ holds for all $\vec{X}$, so that (21) holds for all $\vec{\lambda}$, with $\vec{\lambda}'$ given by (22). Substituting (22) into (21) gives us the identity in $\lambda_0, \lambda_1, \dots, \lambda_{n-1}$:

$$(23) \qquad \prod_{h=0}^{n-1}\left(\sum_{j=0}^{n-1} \alpha_{hj}\lambda_j\right) = \prod_{h=0}^{n-1}\lambda_h.$$

For each h the right side vanishes identically for $\lambda_h = 0$. It follows that the factors of the left side must be of the form $\alpha_h\lambda_h$, where $\prod_{h=0}^{n=1} \alpha_h = 1$, and hence (21) must have the form

$$(24) \qquad \lambda'_{\sigma(h)} = \alpha_h\lambda_h \quad \text{for} \quad 0 \leqslant h \leqslant n-1,$$

where σ is a permutation on $\{0,1,\dots,n-1\}$ and where $\prod_{h=0}^{n-1} \alpha_h = 1$, i.e.,

$$(25) \qquad \vec{\lambda}' = \left(\sum_{h=0}^{n-1} \alpha_h E_{\sigma(h),h}\right)\vec{\lambda}, \quad \text{where} \quad \prod_{h=0}^{n-1} \alpha_h = 1,$$

and where σ is a permutation on $\{0,1,\dots,n-1\}$. Let

$$(26) \qquad A = \sum_{h=0}^{n-1} \alpha_h E_{\sigma(h),h}, \quad \text{where} \quad \prod_{h=0}^{n-1} \alpha_h = 1.$$

We have the equations

$$(27.1) \qquad \vec{X} = B^{-1}\vec{\lambda}, \quad \vec{X}' = B^{-1}\vec{\lambda}'$$

and

$$(27.2) \qquad \vec{\lambda}' = A\vec{\lambda},$$

the latter given by equations (25) and (26). Then

$$\vec{X}' = B^{-1}\vec{\lambda}' = B^{-1}A\vec{\lambda} = B^{-1}AB\vec{X}.$$

So G's leading to $(***)$ must be of the form

$$(28) \qquad G = B^{-1}AB.$$

Substituting equations (14), (15), and (26) into (28) gives us

$$(29) \qquad G = \frac{1}{n}\sum_{h,j}\left(\sum_{q=0}^{n-1} \zeta^{-h\sigma(q)+jq}\alpha_q\right)E_{h,j},$$

where σ is a permutation on $\{0,1,\dots,n-1\}$ and $\prod_{q=0}^{n-1} \alpha_q = 1$. This is equation $(*)$. Q.E.D.

Proof of $(\Rightarrow)$. If G has form $(*)$, then $G = B^{-1}AB$. Now since $\vec{X}' = B^{-1}AB\vec{X}$, it follows that $B^{-1}\vec{\lambda}' = B^{-1}ABB^{-1}\vec{\lambda}$, or $\vec{\lambda}' = A\vec{\lambda}$ where A is given by equation (26). This equation and $\vec{\lambda}' = A\vec{\lambda}$ imply equation (21) which in turn implies condition $(***)$. Q.E.D.

Equation $(*)$ can be written as $G = B^{-1}PDB$, where B and B^{-1} are given by equations (14) and (15), P is a permutation matrix, and D is a diagonal matrix whose determinant equals 1.

The set of transformations $(*)$ form a group. Let p be a permutation on $\{0,1,\dots,n-1\}$ and

let $\beta_0, \beta_1, \ldots, \beta_{n-1} \in \mathbb{C}$ such that $\prod_{h=0}^{n-1} \beta_h = 1$. Let

$$G' = \frac{1}{n} \sum_{h,j} \left(\sum_{q=0}^{n-1} \zeta^{-hp(q)+jq} \beta_q \right) E_{h,j}.$$

Then

$$GG' = \frac{1}{n} \sum_{h,j} \left(\sum_{q=0}^{n-1} \zeta^{-h\sigma p(q)+jq} \alpha_{p(q)} \beta_q \right) E_{h,j}.$$

As an example of all this, if $n = 2$, then

$$B = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad \text{and} \quad B^{-1} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

A can take the form $\pm \begin{pmatrix} e^{\psi} & 0 \\ 0 & e^{-\psi} \end{pmatrix}$ for $\psi \in \mathbb{C}$. Then

$$G = B^{-1}AB = \pm \begin{pmatrix} \operatorname{ch} \psi & \operatorname{sh} \psi \\ \operatorname{sh} \psi & \operatorname{ch} \psi \end{pmatrix}$$

which is plus or minus the Lorentz transformation.

3. Skew-circulant matrices. These are matrices such as

$$\begin{bmatrix} 2 & 5 & 8 & 9 \\ -9 & 2 & 5 & 8 \\ -8 & -9 & 2 & 5 \\ -5 & -8 & -9 & 2 \end{bmatrix}.$$

These matrices are like circulant matrices with minuses on the entries in the lower-triangular part. They have many properties like those of circulant matrices.

The eigenvalues $\lambda_0, \lambda_1, \ldots, \lambda_{n-1}$ of the $n \times n$ complex skew-circulant matrix X with first row $(x_0, x_1, \ldots, x_{n-1})$ have the form

$$\lambda_h = \sum_{j=0}^{n-1} \zeta^{hj} \beta^j x_j \quad \text{for} \quad 0 \leqslant h \leqslant n-1, \quad \text{where} \quad \beta = e^{\pi i/n}, \tag{30}$$

and the corresponding eigenvectors are

$$\vec{W}_h = \operatorname{Col}(1, \zeta^h \beta, \zeta^{2h} \beta^2, \ldots, \zeta^{(n-1)h} \beta^{n-1}) \quad \text{for} \quad 0 \leqslant h \leqslant n-1. \tag{31}$$

The eigenvectors form the matrix

$$W = \sum_{h,j} \zeta^{hj} \beta^h E_{h,j}, \tag{32}$$

which has an inverse

$$W^{-1} = \frac{1}{n} \sum_{h,j} \zeta^{-hj} \beta^{-j} E_{h,j}; \tag{32}$$

and so

$$W \operatorname{diag}(\lambda_0, \lambda_1, \ldots, \lambda_{n-1}) W^{-1} = X. \tag{33}$$

Let, however,

$$V = \sum_{h,j} \zeta^{hj} \beta^j E_{h,j}. \tag{34}$$

The equations (30) take the form

$$\vec{\lambda} = V\vec{X}.$$

Also

$$V^{-1} = \frac{1}{n}\sum_{h,j} \zeta^{-hj}\beta^{-h}E_{h,j}. \tag{35}$$

Since the determinant of a matrix is the product of its eigenvalues, by Davis ([1], p. 84),

$$\det X = \prod_{h=0}^{n-1}\left(\sum_{j=0}^{n-1} \zeta^{hj}\beta^{j}x_j\right). \tag{36}$$

We can restate the Theorem with the term skew-circulant matrix instead of circulant matrix and with $(*)$ replaced by

$$G = \frac{1}{n}\sum_{h,j}\left(\sum_{q=0}^{n-1} \zeta^{-h\sigma(q)+jq}\beta^{-h+j}\alpha_q\right)E_{h,j}, \tag{$*'$}$$

and get a true statement. In proof, we can show that $G = V^{-1}AV$, where

$$A = \sum_{h=0}^{n-1} \alpha_h E_{\sigma(h),h} \quad \text{such that} \quad \prod_{h=0}^{n-1}\alpha_h = 1, \tag{37}$$

and σ is a permutation on $\{0,1,\ldots,n-1\}$. We get then equation $(*')$. The proof of the converse is like that for the Theorem.

References

1. Philip J. Davis, Circulant Matrices, Wiley-Interscience, New York, 1974.

2. Thomas Muir, The Theory of Circulants in the Historical Development up to 1860, Proceedings of the Royal Society of Edinburgh, Vol. 26, November 1905 to July 1906, pp. 390–398.

APPROXIMATING POLYGONS FOR LEBESGUE'S AND SCHOENBERG'S SPACE FILLING CURVES

HANS SAGAN

Department of Mathematics, North Carolina State University, Raleigh, NC 27650

Background. In the olden days, a curve was considered to be a geometric object that is generated by a continuously moving point during a finite and closed time interval, or as we would put it today, the continuous image (in E^n) of a line segment. In 1890, G. Peano [8] shook differential geometers out of their complacency by demonstrating that this definition embraces objects of positive (two- or higher- dimensional) Jordan content (such as squares and cubes) that nobody would wish to call curves. The following year, D. Hilbert [3] modified Peano's example and replaced Peano's arithmetical definition by a very simple and compelling geometric generating procedure:

Since he received his Ph.D. degree in Mathematics under J. Radon from the University of Vienna in 1950, Hans Sagan, who speaks English and German fluently, French haltingly, Latin reluctantly, and has a smattering of classical Greek, has taught in three countries on two continents. He has also lectured widely at some 30 Colleges and Universities in fifteen States and three Canadian Provinces and is the author or co-author of eight books. As an ardent sailor with sailing experience on five bodies of water, including the Atlantic, he always had a soft spot for jagged curves, generalized or space-filling, no matter.

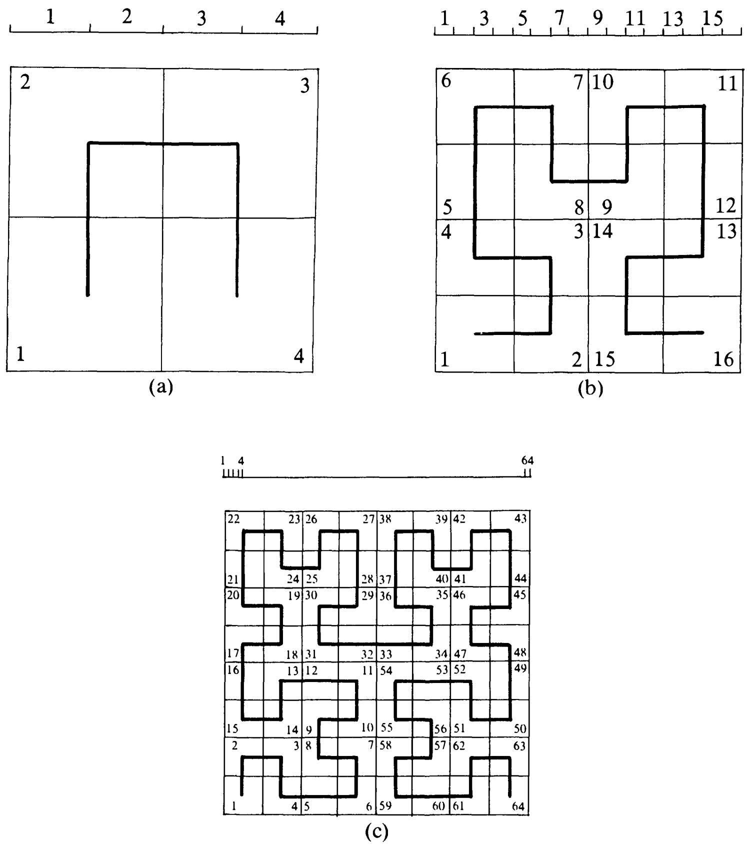


FIG. 1. Generating the space-filling curve by Hilbert.

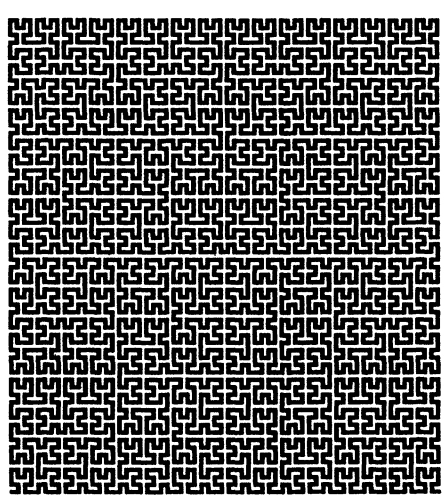


FIG. 1(d). Sixth approximating polygon for Hilbert's space-filling curve.

One partitions the interval $[0,1]$ into 4 congruent intervals and the square $[0,1] \times [0,1]$ into 4 congruent squares which are to be enumerated so that each two consecutively numbered squares have an edge in common. (See Fig. 1(a).) If the entire interval can be mapped onto the entire square, then $[0, \frac{1}{4}]$ can be mapped onto square 1, $[\frac{1}{4}, \frac{1}{2}]$ can be mapped onto square 2, etc. Again, partition every subinterval into 4 congruent intervals, each sub-square into 4 congruent squares, enumerate the squares so that the preceding mapping is preserved and that each two consecutive squares have an edge in common (see Fig. 1(b)) and repeat the previous argument. The next step is depicted in Fig. 1(c). Proceed in this manner ad infinitum to obtain a mapping that is, by construction, continuous. That it is also surjective follows from the fact that each point in $[0,1] \times [0,1]$ lies in a sequence of closed nested squares the diameters of which shrink to zero and which corresponds to a sequence of closed nested intervals the lengths of which shrink to zero also.

The sequence in which the squares are to be taken may be symbolized by a polygonal path through the midpoints of the squares in the order of their enumeration. W. Wunderlich **[10]** calls such polygons *approximating polygons*. The approximating polygon, corresponding to the 6th step, is depicted in Fig. 1(d). (This slightly distorted figure was generated on the IBM PC, using a program by Eui In Lee **[6]**.) We will adopt a related notion of approximating polygons in the next two sections.

In general, one calls

$$\left.\begin{aligned} x &= f(t) \\ y &= g(t) \end{aligned}\right\} \quad a \leqslant t \leqslant b \tag{1}$$

a two-dimensional *space-filling curve* or a *Peano-curve* if f, g are continuous and the set $\{(f(t), g(t)) \in E^2 | a \leqslant t \leqslant b\}$ has positive 2-dimensional Jordan content. Incidentally, such a mapping cannot possibly be bi-jective because any bi-jective map from a line onto a surface is, by necessity, discontinuous, as E. Netto has already shown **[7]**.

Following Peano's and Hilbert's examples, many other space-filling curves have been proposed. Most are modifications of Hilbert's example by partitioning the interval and square into 9 parts at each step, or by mapping onto an isosceles right triangle, or such (see for example **[4]**). Two, however, do not fit into this group, namely the ones by H. Lebesgue and I. J. Schoenberg.

Lebesgue's space-filling curve. H. Lebesgue **[5]** defines f, g in (1) with $a = 0, b = 1$ as follows: For all members of the Cantor set

$$t_0 = \frac{2a_1}{3} + \frac{2a_2}{3^2} + \frac{2a_3}{3^3} + \frac{2a_4}{3^4} + \frac{2a_5}{3^5} + \frac{2a_6}{3^6} + \cdots, \quad a_i = 0 \text{ or } 1, \tag{2}$$

let

$$f(t_0) = \frac{a_1}{2} + \frac{a_3}{2^2} + \frac{a_5}{2^3} + \cdots, \quad g(t_0) = \frac{a_2}{2} + \frac{a_4}{2^2} + \frac{a_6}{2^3} + \cdots. \tag{3}$$

Since

$$|f(t) - f(t_0)| < \frac{1}{2^n} \text{ for all } t = \frac{2a_1}{3} + \frac{2a_2}{3^2} + \cdots + \frac{2a_{2n}}{3^{2n}} + \frac{2\alpha_{2n+1}}{3^{2n+1}}$$
$$\frac{2\alpha_{2n+2}}{3^{2n+2}} + \cdots, \quad \alpha_i = 0 \text{ or } 1,$$

and a similar result obtains for g, we see that f, g are continuous on the Cantor set. We now extend the definition of f, g to the complement of the Cantor set (which is a denumerable union of open intervals) continuously by linear interpolation. For example, to extend the definition of f, g to the interval $(1/3, 2/3)$, we note that

$$f\left(\frac{1}{3}\right) = \frac{1}{2}, f\left(\frac{2}{3}\right) = \frac{1}{2}, g\left(\frac{1}{3}\right) = 1, g\left(\frac{2}{3}\right) = 0$$

and set accordingly $f(t) = 1/2$, $g(t) = -3t + 2$, for $1/3 < t < 2/3$, and continue in this manner with the intervals $(1/9, 2/9)$, $(7/9, 8/9)$, etc. By construction, the extended functions are continuous on the complement of the Cantor set. To show that they are continuous on the entire interval, one only has to note that every point of the Cantor set that is not a left endpoint of one of the open intervals in the complement is a left accumulation point of the Cantor set and, by the same token, that every point in the Cantor set that is not a right endpoint of one of those intervals is a right accumulation point of the Cantor set. For the gory details, see [2]. That the mapping is surjective may be seen as follows: If $(x, y) \in [0,1] \times [0,1]$, then these coordinates may be represented as in (3) and the point appears, by construction, as the image of t_0 as in (2).

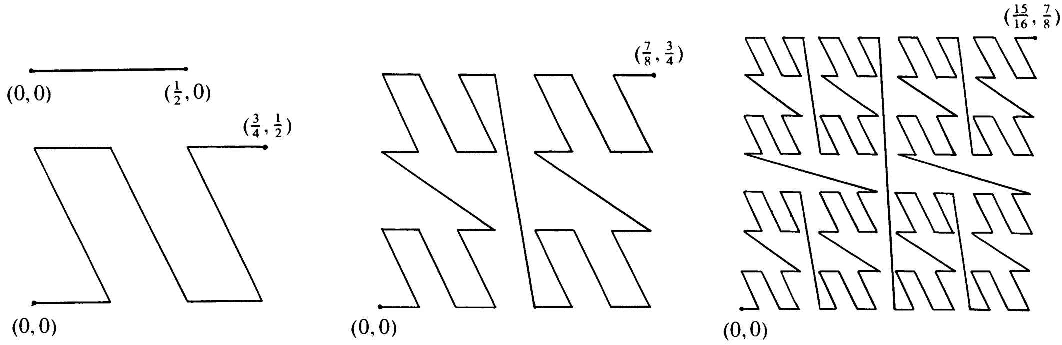


FIG. 2. Lebesgue polygons of order 1, 3, 5, 7.

It is reasonable to call a polygon with vertices for

$$t_m = \frac{2a_1}{3} + \frac{2a_2}{3^2} + \cdots + \frac{2a_m}{3^m}, \quad a_i = 0 \text{ or } 1,$$

for fixed m, an approximating polygon of mth order for Lebesgue's space-filling curve or, for short, a *Lebesgue-polygon of mth order.* Sketches of such polygons convey some idea of how the square will eventually be filled. For example, to obtain the Lebesgue polygon of 2nd order, we consider the parameter values (in ternary representation) 0.00, 0.02, 0.20, 0.22 to obtain the vertices $(0,0)$, $(0,1/2)$, $(1/2,0)$, $(1/2,1/2)$. The Lebesgue polygons of order 1, 3, 5, 7 are to be found in Fig. 2 and the ones of order 2, 4, 6 in Fig. 3. Observe how these polygons meander all over the square to form attractive geometric patterns. There is an obvious strategy for drawing

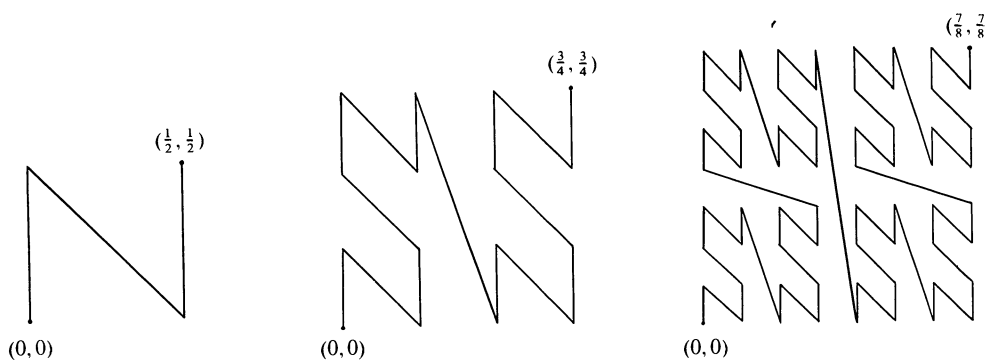


FIG. 3. Lebesgue polygons of order 2, 4, 6.

these polygons. To obtain the polygon of order $m + 2$, copy the one of mth order 4 times in a square pattern, connect the 4 parts suitably by joins, and shrink the result to one half of its original size.

The joins of various lengths approximate the ultimate vertical or horizontal joins that correspond to the linear extensions of f, g into the intervals $(1/3, 2/3)$, $(1/9, 2/9)$, $(7/9, 8/9)$, etc. For example, the deep plunge in the middle is an approximation for the ultimate vertical plunge from $(1/2, 1)$ to $(1/2, 0)$ corresponding to the linear extensions of f, g into $(1/3, 2/3)$. Note that f, g are differentiable a.e. as opposed to the functions that generate a Hilbert-type "curve".

Schoenberg's space-filling curve. Undoubtedly inspired by Lebesgue's example, I. J. Schoenberg published his version of a space-filling curve in 1938 **[9]**. His curve, incidentally, coincides with Lebesgue's curve on the Cantor set. No drawings of approximating polygons appear to exist—and for good reasons as we will see.

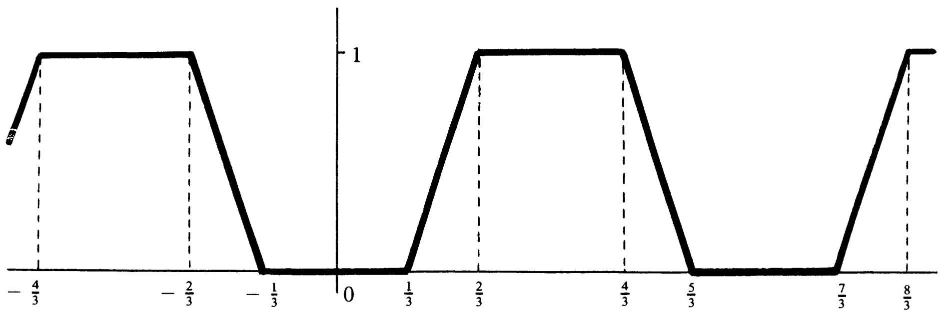


FIG. 4. Graph of generating function p.

Schoenberg's curve is defined in terms of the even 2-periodic generating function

$$p(t) = \begin{cases} 0 & \text{for } 0 \leqslant t < 1/3 \\ 3t - 1 & \text{for } 1/3 \leqslant t < 2/3 \\ 1 & \text{for } 2/3 \leqslant t \leqslant 1 \end{cases} \quad p(-t) = p(t), p(t+2) = p(t)$$

(see Fig. 4) as follows:

$$f(t) = \sum_{k=1}^{\infty} p(3^{2k-2}t)/2^k, \quad g(t) = \sum_{k=1}^{\infty} p(3^{2k-1}t)/2^k, \tag{4}$$

with f, g as in (1) and $a = 0$, $b = 1$. The proof that this curve is indeed space-filling is of such stunning simplicity that it has become standard Advanced Calculus fare. Clearly, f, g, being represented by uniformly convergent series of continuous functions, are continuous. Since any point in $[0, 1] \times [0, 1]$ may be represented as in (3), we see from (4) that such a point is the image of t_0 as in (2), demonstrating that the mapping is surjective and, at the same time, that Schoenberg's curve coincides with Lebesgue's curve on the Cantor set. For more details, see **[1]**. f, g are nowhere differentiable.

From (4), Schoenberg's curve has potential vertices for $t_{m,n} = n/3^m$, for $m = 1, 2, 3, \ldots, n = 0, 1, 2, \ldots, 3^m$. In conformity with our previous definition, we call the polygon with vertices for $t_{m,n}$, for fixed $m, n = 0, 1, \ldots, 3^m - 1$, an approximating polygon of mth order for Schoenberg's curve or, simply, a *Schoenberg polygon of mth order*. It is a simple matter to draw the Schoenberg polygons of 1st and 2nd order and it is still within the capabilities of a reasonably patient person to draw the ones of 3rd and 4th order. (See Fig. 5 and Fig. 6.) To draw the one of 5th order with 243 potential vertices and no simple pattern to serve as a guide would tax the endurance and

manipulative skills of a 17th century mathematician, and the mere thought of going beyond that boggles the mind.

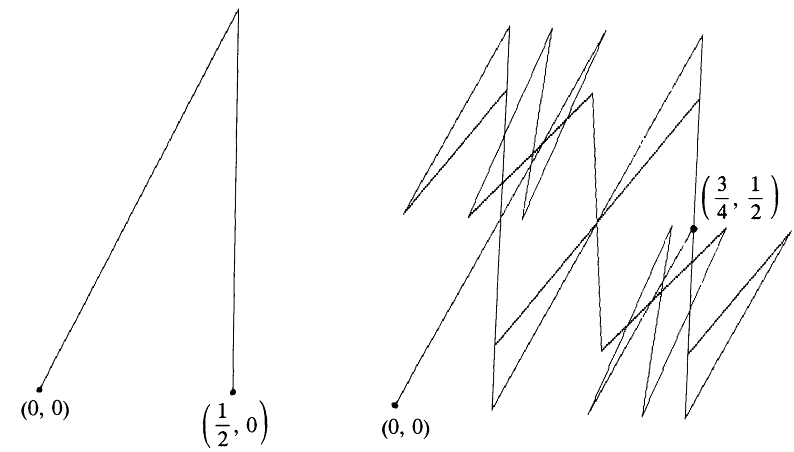


FIG. 5. Schoenberg polygons of order 1, 3.

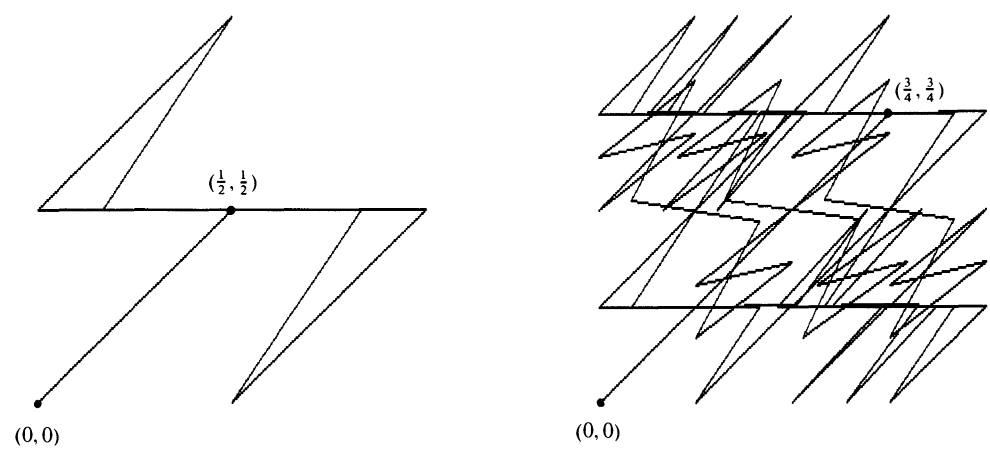


FIG. 6. Schoenberg polygons of order 2, 4.

To obtain the Schoenberg polygons of higher order, we have written a simple program (in BASIC) for the IBM PC that is capable, in principle, of producing graphs of the Schoenberg polygons of any order, were it not for the built-in inadequacies of the computer such as its limited precision in the arithmetical operations and the restriction imposed by a 640 by 200 pixels high-resolution screen. We may avoid infinite sums when evaluating the coordinates of the vertices, if we note that

$$p(3^j n/3^m) = p(n) = n - 2[n/2] \quad \text{for all } j \geqslant m.$$

Hence we have

$$f(t_{m,n}) = \sum_{k=1}^{[(m+1)/2]} p\left(\frac{3^{2k-2}n}{3^m}\right)/2^k + \frac{n - 2[n/2]}{2^{[(m+1)/2]}},$$

$$g(t_{m,n}) = \sum_{k=1}^{[(m+1)/2]} p\left(\frac{3^{2k-1}n}{3^m}\right)/2^k + \frac{n - 2[n/2]}{2^{[(m+1)/2]}}.$$

```
5 REM This program produces the graph of the approximating polygon of m-th
      order for Schoenberg's Peano-Curve. The origin is placed at the left
      margin, 199 pixels down, at (0,198).
10 CLS:PRINT"m = ";:INPUT M:SCREEN 2:PSET (0,198):N=1
15 REM The generating function p is defined in lines 20 - 50 by making use of
       the unit-step function u(t,a) = sgn(sgn(t-a)+1).
20 DEF FNA(T)=(3*T-1)*(SGN(SGN(3*T-1)+1)-SGN(SGN(3*T-2)+1))
30 DEF FNB(T)=SGN(SGN(3*T-2)+1)-SGN(SGN(3*T-4)+1)
40 DEF FNC(T)=(5-3*T)*(SGN(SGN(3*T-4)+1)-SGN(SGN(3*T-5)+1))
50 DEF FND(T)=FNA(T)+FNB(T)+FNC(T)
60 X=0:Y=0:T=N/(3^M):L=1
65 REM In lines 70 and 90, we reduce t modulo 2 by t = t - 2*int(t/2).
       Potential vertices occur for t = n/(3^m) for n = 0,1,2,...,3^m-1.
       x,y are evaluated for each potential vertex in lines 70 - 100.
70 X=X+FND(T)/(2^L):T=3*T:T=T-2*INT(T/2)
80 Y=Y+FND(T)/(2^L):L=L+1:IF L>INT((M+1)/2) THEN 100
90 T=3*T:T=T-2*INT(T/2):GOTO 70
100 X=X+(N-2*INT(N/2))/(2^(L-1)):Y=Y+(N-2*INT(N/2))/(2^(L-1))
105 REM Each vertex is joined to its predecessor by a straight line in 110.
110 LINE -(480*X,198-198*Y),1
120 N=N+1:IF N>3^M-1 THEN 140
130 GOTO 60
140 IF INKEY$="" THEN 140
150 END
```

FIG. 7. Program for drawing Schoenberg polygons.

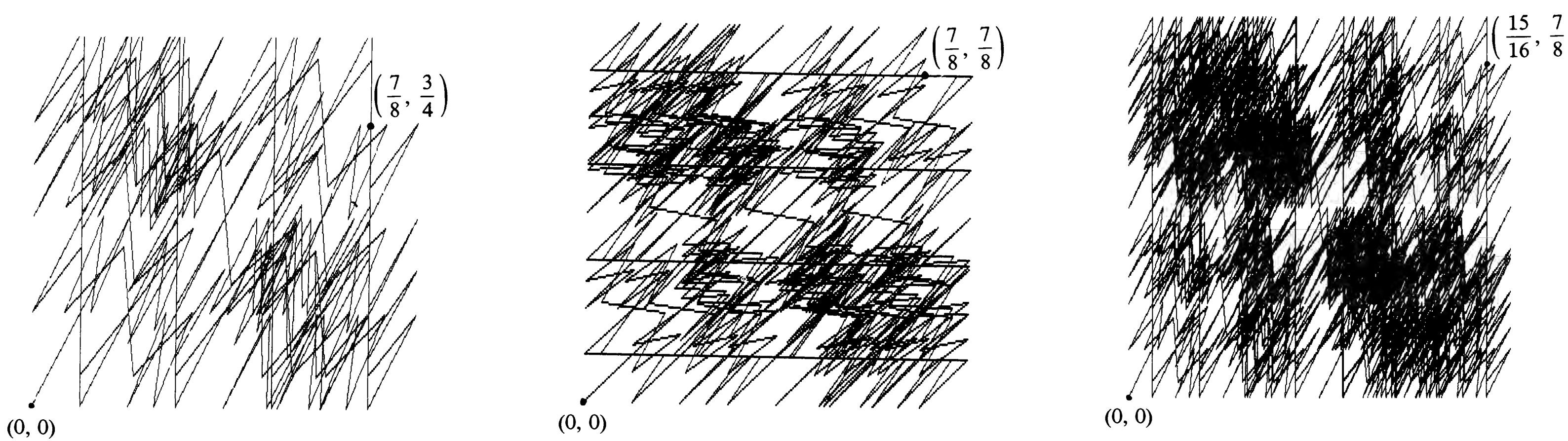


FIG. 8. Schoenberg polygons of order 5, 6, 7.

The program is displayed in Fig. 7. The Schoenberg polygons of 5th, 6th, and 7th order are displayed in Fig. 8.

There is a discernible pattern, albeit a very complicated one. Attractive it is not. It would appear that what Arnold Schoenberg has done to music, I. J. Schoenberg has done to Peano-curves. Let it be a challenge to the geometers to come up with a strategy for the construction of Schoenberg polygons. If the reader wishes to run the program in Fig. 7, he is advised to compile it first (without the REMs) by means of the BASIC compiler that goes with the IBM PC to achieve a substantial reduction of the execution time. Since the Schoenberg curve and the Lebesgue curve coincide on the Cantor set, the program in Fig. 7 will also produce the Lebesgue polygons if T, in line 60, is restricted to the appropriate members of the Cantor set. With double precision, one can achieve a serviceable image of the Lebesgue polygon of 13th order. A final note: Had we adopted the same notion of approximating polygon to the Hilbert curve, polygons other than the ones in Fig. 1 would have emerged. However, as the sub-squares shrink, their midpoints and the images of the beginning points of the subintervals move closer together and, in the limit, the one will approach the other.

References

1. T. Apostol, Mathematical Analysis, 2nd ed., Addison-Wesley, Reading, MA, 1974, p. 224.
2. A. Devinatz, Advanced Calculus, Holt, Rinehart, Winston, New York, 1968, p. 253.
3. D. Hilbert, Ueber die stetige Abbildung einer Linie auf ein Flaechenstueck, Math. Ann., 38 (1891) 459–460.
4. K. Knopp, Einheitliche Erzeugung und Darstellung der Kurven von Peano, Osgood und von Koch, Arch. Math. Phys., 26 (1917) 103–115.
5. H. Lebesgue, Leçons sur l'Intégration et la Recherche des Fonctions Primitives, Gauthier-Villars, Paris, 1904, pp. 44–45.
6. Eui In Lee, Recursive Curves, PC Tech J. 3, No. 4 (April 1985) 171–174.
7. E. Netto, Beitrag zur Mannigfaltigkeitslehre, Crelle J., 86 (1879) 263.
8. G. Peano, Sur une courbe qui remplit toute une aire plane, Math. Ann., 36 (1890) 157–160.
9. I. J. Schoenberg, The Peano-Curve of Lebesgue, Bull. Amer. Math. Soc., 44 (1938) 519.
10. W. Wunderlich, Ueber Peano-Kurven, Elem. Math., 28 (1973) 1–10.

PROGRESS REPORTS

EDITED BY THOMAS BANCHOFF AND RICHARD MILLMAN

It is easy to be too busy to pay attention to what anyone else is doing, but not good. All of us should know, and want to know, what has been discovered since our formal education ended, but new words, and relations between them, are growing too fast to keep up. It is possible for a person to learn of the title of a recent work and of the key words used in it and still not have the faintest idea of what the subject is.

Progress Reports is to be an almost periodic column intended to increase everyone's mathematical information about what others have been up to. Each column will report one step forward in the mathematics of our time. The purpose is to inform, more than to instruct: what is the name of the subject, what are some of the words it uses, what is a typical question, what is the answer, who found it. The emphasis will be on concrete questions and answers (theorems), and not on general contexts and techniques (theories). References will be kept minimal: usually they will include only one of the earliest papers in which the answer appears and a more recent exposition of the discovery, whenever one is easily available.

Everyone is invited to nominate subjects to be reported on and authors to prepare the reports. The ground rules are that the principal theorem should be old enough to have been published in the usual sense of that word (and not just circulated by word of mouth or in preprints); it should be of interest to more than just a few specialists; and it should be new enough to have an effect on the mathematical life of the present and near future. In practice most reports will probably be on progress achieved somewhere between 5 and 15 years ago.

MACDONALD'S η-FUNCTION FORMULA AND SOME DEVELOPMENTS IN DIFFERENTIAL GEOMETRY

HOWARD D. FEGAN
Department of Mathematics, University of New Mexico, Albuquerque, NM 87131

RICHARD S. MILLMAN
Division of Mathematical Sciences, National Science Foundation, Washington, DC 20550

Who amongst us can resist the beauty of a theory that unifies, explains, simplifies and expands our mathematical knowledge, even if this might cost additional sophistication? Since the nineteenth century, mathematicians have obtained many fascinating and mysterious formulae in the theory of special functions. Consider for instance, the theta function,

$$\theta(t) = 1 + 2\sum \exp(i\pi n^2 t).$$

It satisfies the identity $\theta(t) = \theta(t+2)$ and $\theta(-1/t) = \sqrt{t/i}\,\theta(t)$ but what do these really mean? Another example deals with the Euler phi function, $\phi(x) = \Pi_n(1 - x^n)$ which is intimately connected to number theory and combinatorics in the following way. Recall that the partition function, $p(n)$, is the number of ways that n can be written as the sum of positive integers. It is a classical fact that

$$\phi(x)^{-1} = \sum p(n)x^n,$$

hence $\phi(x)$ is related to combinatorics. One more example is the expansion of $\phi(x)^{10}$ due to Winquist which leads to a simple proof of the combinatorial result $p(11m + 6) \equiv 0 \bmod 11$. The connection to number theory is best illustrated by the τ-function of Ramanujan. This function is defined by the expansion

$$\sum_{n=1}^{\infty} \tau(n)x^n = x\phi(x)^{24},$$

and one has the elegant formula:

$$\tau(n) = \sum \frac{(a-b)(a-c)(a-d)(a-e)(b-c)(b-d)(b-e)(c-d)(c-e)(d-e)}{1!\,2!\,3!\,4!},$$

summed over all sets of integers a, b, c, d, e with

$$a, b, c, d, e \equiv 1, 2, 3, 4, 5 \bmod 5,$$
$$a + b + c + d + e = 0 \quad \text{and} \quad a^2 + b^2 + c^2 + d^2 + e^2 = 10n.$$

This relationship is due to Dyson, see **[1]**, and is related to an expansion of $\phi(x)^{24}$. What is the meaning of the strange invariance properties illustrated by the ϕ function or the bizarre exponents like 10 or 24?

Even with hindsight it is difficult to see what the unifying theme is in the above paragraph or even if such a thing should exist. The nexus is the Laplace operator and its eigenvalues as is illustrated by a specific case. Consider the one-dimensional Laplace operator, $\Delta = -d^2/dx^2$, operating on functions which are periodic with period 2π. The eigenvalues of Δ are 0 (with multiplicity 1) and n^2 (with multiplicity 2) for each positive integer, n. The first observation is that the theta function is exactly $\Sigma \exp(-\lambda t)$, where the sum goes over all eigenvalues, λ, of Δ. (Truthfully, it isn't exact: there is a slight problem with the normalization $i\pi$. Remember, the multiplicities count in the sum.) The second observation is that considering the Laplace operator on the space of periodic functions is really looking at Δ as an operator on functions on a circle of unit radius. A Lie group is, roughly speaking, a subgroup of the group of nonsingular matrices $Gl(n)$ and on which one may do calculus as it is a subset of Euclidean space. As the circle is a Lie

group (it is $SO(2) \subset Gl(2)$), a reasonable direction for generalization and possible unification is now possible: see whether there is an analogy of the Laplace operator and the theta type expansion on other Lie groups. Macdonald's work of 1972 extends this miracle from the circle to other compact Lie groups. In fact the number 10 and 24 are the dimensions of certain compact Lie groups.

A compact Lie group determines and is determined (essentially) by a root system. A root system is a special set of vectors in a finite dimensional vector space. Joining another special vector (called the "negative of the highest root") to a root system gives a set which generates an affine root system. Using combinatorial techniques Macdonald obtained an analogous formula to Jacobi's triple product from each affine root system. These are called η-function identities because the Dedekind η-function is

$$\eta(t) = e^{i\pi t/12}\phi(e^{2\pi i t}).$$

This function is of interest to number theorists. (It is something they call "a modular form". So is θ which is what the first identities for θ say.) Clearly any expression for $\phi(x)^d$ is easily translated into one for $\eta(t)^d$. The reader is referred to [3] for the details of these results.

On a compact Lie group there is a Laplacian. Hence there is the heat equation, disguised by using complex time:

$$\Delta u - (1/2\pi i)\,\partial u/\partial t = 0.$$

Kostant showed that there was an infinite product solution to this equation. This was done by starting with Macdonald's identity and rewriting the sum in terms of a Fourier series on the Lie group. The main work was essentially in identifying the initial data for such a solution. This data is a delta distribution concentrated along an orbit of a special point (called "principal of type ρ") in the group.

To illustrate these results we give the case of the group $SU(2)$ whose dimension $d = 3$. These results were first obtained by Jacobi from his work on theta functions and published in 1829. The classical theta function is defined by

$$\theta_1(z, t) = 2\sum_{n=0}^{\infty}(-1)^n \sin\{(2n+1)z\}\, e^{i\pi(n+1/2)^2 t}.$$

Jacobi's triple product identity is

$$\theta_1(z, t) = 2\phi(e^{2\pi i t})\, e^{i\pi t/4}\sin z \prod_{n=1}^{\infty}(1 - 2e^{2i\pi n t}\cos 2z + e^{4i\pi n t}).$$

Equating these two expressions for θ_1, writing x for $e^{2i\pi t}$ and setting $z = 0$ gives rise to:

$$\phi(x)^3 = \sum_{n=0}^{\infty}(-1)^n(2n+1)\,x^{n(n+1)/2}. \tag{1}$$

Notice that we use L'Hôpital's rule to evaluate $\lim_{z\to 0}\sin((2n+1)z)/\sin z = 2n+1$. The theta function satisfies the differential equation

$$\frac{\partial^2\theta_1}{\partial z^2} + \frac{4}{i\pi}\frac{\partial\theta_1}{\partial t} = 0.$$

This equation looks like the Schroedinger equation but is really the heat equation in disguise. To see through the disguise one recalls that t is complex with strictly positive imaginary part. For example if $t = i\tau$ is pure imaginary, then a simple change of variables turns the equation into

$$\frac{\partial^2\theta_1}{\partial z^2} - \frac{4}{\pi}\frac{\partial\theta_1}{\partial \tau} = 0.$$

For other t we extend this equation by analytic continuation.

Another development from Macdonald's work has been algebraic. Just as a root system corresponds to a Lie algebra, and hence to a Lie group, so an affine root system corresponds to an infinite dimensional Lie algebra. The generalization of these are called Kac-Moody Lie algebras and analogous identities can be obtained from these algebras. This generalization has relevance to physics in the topic of super gravity and was one subject, amongst others, of several recent conferences. One such conference was the AMS summer conference on Mathematics and Physics held in Chicago in 1982.

For differential geometry the most surprising aspect of this work has been the discovery of a new and interesting direction in the theory of compact Lie groups. For many years compact Lie groups have been used in mathematics and physics. There is now an unexpected link between differential geometry and number theory. Kostant showed the relevance to differential geometry in **[2]** as a significant development from Macdonald's work. Recently Macdonald has given a generalization of his results, see **[3]**. As a result of these identities we have realized that we do not know everything about these groups and work has started on reviving their study.

References

1. F. J. Dyson, Missed opportunities, Bull. Amer. Math. Soc., 78 (1972) 635–652.

2. B. Kostant, On Macdonald's η-function formula, the Laplacian and generalized exponents, Adv. in Math., 20 (1976) 179–212.

3. I. G. Macdonald, Some conjectures in root systems, SIAM J. Math Anal., 13 (1982) 988–1007.

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics. The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

WANTED: THE GENERAL SOLUTION OR ELSE

IH-CHING HSU

Microwave & Support Systems Division, Sperry Corporation, Clearwater, FL 33518

In this note we propose a class of functional equations and pose three related problems of varying generality. The first problem concerns the following class of functional equations:

$$F(x, y) + F(\phi(x), \psi(y)) = F(x, \psi(y)) + F(\phi(x), y), \tag{1}$$

where ϕ and ψ are given functions, and we seek a general solution F. The second problem involves the same formal functional equation (1) but the given functions ϕ and ψ are now assumed to be involutions; that is $\phi \circ \phi = \psi \circ \psi =$ identity. The third problem is a further specialization to $\phi(t) = \psi(t) = 1 - t$; namely, we seek the general solution of

$$F(x, y) + F(1 - x, 1 - y) = F(x, 1 - y) + F(1 - x, y). \tag{2}$$

In each of these problems, appropriate domains are assumed for the given functions ϕ and ψ, and for the subject function F.

Several remarks are in order: (a) The solution set to each of the functional equations is a linear space. (b) Whenever $F(x, y)$ is a solution to the second problem, $F(\phi(x), y)$ and $F(x, \psi(y))$, and hence $F(\phi(x), \psi(y))$, are solutions. (c) For arbitrary f and g, $F(x, y) = f(x) + g(y)$ is a solution to all three problems. However, this function F is not the general solution to the equations in the second or third problems. The following additional remark and theorem offer some other solutions.

REMARK. Under the transformation $F = f \circ H$, with $H(x, y) = xy$ and $f\colon R \to R$, equation (2) reduces to the following form:

$$f(xy) + f\{(1-x)(1-y)\} = f\{x(1-y)\} + f\{(1-x)y\}. \tag{3}$$

Equation (3) appears in [1] as a problem proposed by K. Lajko. In [2], C. J. Eliezer, employing elementary methods, establishes f_1 and f_2, where $f_1(x) = 1$ and $f_2(x) = x^2 - x$, as a basis for the subspace of differentiable solutions to equation (3). The referee observes that with $F = f_2 \circ H$,

$$F(x, y) = f_2(xy) = (xy)^2 - xy$$

is readily established for equation (2) as a solution, one not in the form of $f(x) + g(y)$.

The idea of seeking solutions in composite function form is carried over to the following theorem.

THEOREM. *For each arbitrary function* $G\colon R^2 \to R$, *with* $H\colon R^2 \to R^2$ *defined by* $H(x, y) = (\phi(x) + x, \psi(y) + y)$, $F = G \circ H$ *is a solution to equation* (1) *with* ϕ *and* ψ *given and* $\phi \circ \phi = \psi \circ \psi =$ *identity.*

Proof. $F(x, y) + F(\phi(x), \psi(y)) = G(\phi(x) + x, \psi(y) + y) + G(x + \phi(x), y + \psi(y)) = F(x, \psi(y)) + F(\phi(x), y)$.

To conclude, we note that the general solution to each of the functional equations posed here remains unknown. In the absence of a general solution, it would be of interest to characterize the general continuous, or general differentiable solution. Interested readers are referred to [3] [4] for treatments of the general subject.

Acknowledgement. The author wishes to express his thanks to the editor, to the referee, and to Professor Michael Aissen for their comments.

References

1. K. Lajko, Problem, P116, Aequationes Math., 10 (1974) 311.
2. C. J. Eliezer, A solution to $f\{(1-x)(1-y)\} + f(xy) = f\{x(1-y)\} + f\{y(1-x)\}$, Aequationes Math., 10 (1974) 311–312.
3. J. Aczél, Lectures on Functional Equations and their Applications, Academic Press, New York-London, 1966.
4. M. Kuczma, Functional Equations in a Single Variable, Polish Scientific Publishers, Warszawa, 1968.

---

167. MISCELLANEA

Who Defines Whom?

In California, Bill Honig, the Superintendent of Public Instruction, said he thought the general public should have a voice in defining what an excellent teacher should know. "I would not leave the definition of math," Dr. Honig said, "up to the mathematicians."

—The New York Times, October 22, 1985.

A man, a book, a building. (See p. 379.)

$$Ee^{t(X-n)/\sqrt{n}} = e^{-t\sqrt{n}} \sum_{x=0}^{\infty} e^{tx/\sqrt{n}} n^x \frac{e^{-n}}{x!}$$

$$= e^{-t/\sqrt{n}} \exp\{ n(e^{t/\sqrt{n}} - 1)\} = \exp\left\{ \frac{t^2}{2} + O(1/\sqrt{n}) \right\},$$

and Stirling's theorem (1) now follows in exactly the same way as in Section 1.

NOTE. If this proof is tried with n not an integer, it gives the same result but with n replaced by $[n]$, the greatest integer in n.

3. A New Proof. This proof is shorter than the preceding two, is slightly more elementary, and is valid for arbitrary n:

If Z has continuous probability density function $f(\ )$ and integrable characteristic function $\phi(\ )$, then the inversion theorem for characteristic functions, Chung [**1**], p. 143, gives

$$f(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi(t)\, dt. \tag{6}$$

For the Z of Section 1 we have

$$f(0) = \frac{n^{n-.5}e^{-n}}{\Gamma(n)}$$

and, just as done in Section 1 for the moment generating function,

$$\phi(t) = e^{-it\sqrt{n}}(1 - it/\sqrt{n})^{-n} = e^{-t^2/2 + O(1/\sqrt{n})}.$$

Moreover, $\phi(t)$ is easily seen to be dominated by $1/(1 + t^2/2)$ for $n \geq 2$, so the dominated convergence theorem gives

$$\lim_{n\to\infty} \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi(t)\, dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-t^2/2}\, dt = 1/\sqrt{2\pi}.$$

Taking limits of the two sides of (6) for this Z therefore gives

$$\lim_{n\to\infty} \frac{n^{n-.5}e^{-n}}{\Gamma(n)} = 1/\sqrt{2\pi},$$

and multiplication by $\sqrt{2\pi}$ now gives Stirling's Theorem (1).

NOTE. It would be nice to have a proof of the extended Stirling Theorem (2) that is comparable in ease to these proofs of (1). The proof of this section, by using an expansion of $\phi(t)$ in (6), can be used to get a proof of (2), but the proof as we have it is too long and complicated to be satisfactory.

References

1. K. L. Chung, A Course in Probability Theory, Harcourt, Brace & World, New York, 1968.
2. R. A. Khan, A probabilistic proof of Stirling's formula, this MONTHLY, 81 (1974) 366–369.
3. M. Loève, Probability Theory, 3rd ed., Van Nostrand, Princeton, 1963.
4. E. C. Titchmarsh, The Theory of Functions, 2nd ed., Oxford University Press, London, 1939.
5. C. S. Wong, A note on the Central Limit Theorem, this MONTHLY, 84 (1977) 472.

ANSWER TO PHOTO ON PAGE 373

Richard Courant.

NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.
For instructions about submitting Notes for publication in this department see the inside front cover.

A SIMPLE PROOF OF THE IRRATIONALITY OF π^4

JAROSLAV HANČL
Department of Mathematics, Charles University, Prague, Czechoslovakia

It is well known that the number π is transcendental; the proof is complicated. However, there are several elementary proofs that π and π^2 are irrational; see for example [**1**]. Using a simple trick, we will easily show that π^4 is irrational, which of course implies that π and π^2 are also irrational.

THEOREM. *π^4 is irrational.*

Proof. Suppose π^4 is rational. Then there are positive integers r and s such that $\pi^4 = 4r/s$. Let $a = \pi(1+i)/2$, let p denote an integer of the form $1 + 4jrs(j = 1, 2, \dots)$, and let

$$I_p = a^{4p-3} s^p \int_0^1 \frac{(x^4-1)^{p-1}}{(p-1)!} \left(e^{a(x-1)} - e^{-a(x-1)} + ie^{ia(x-1)} - ie^{-ia(x-1)} \right) dx. \tag{1}$$

Repeated integration by parts yields

$$\begin{aligned} I_p = {} & s^p \sum_{k=0}^{4p-4} (-1)^k Q^{(k)}(1) a^{4p-k-4} \left(1 + (-1)^k + (-i)^k + i^k\right) \\ & - s^p \sum_{k=0}^{4p-4} (-1)^k Q^{(k)}(0) a^{4p-k-4} \left(e^{-a} + (-1)^k e^a + (-i)^k e^{-ia} + i^k e^{ia}\right), \end{aligned} \tag{2}$$

where

$$Q(x) = \frac{(x^4-1)^{p-1}}{(p-1)!}.$$

The polynomial $Q(x)$ consists of powers divisible by 4. So $Q^{(4k+s)}(0) = 0$ for every natural number k and $s = 1, 2, 3$. For k divisible by 4 we have

$$e^{-a} + (-1)^k e^a + (-i)^k e^{-ia} + i^k e^{ia} = e^{-a} + e^a + e^{-ia} + e^{ia} = 0.$$

Thus the second sum of (2) is zero.

For k not divisible by 4 we have

$$1 + (-1)^k + i^k + (-i)^k = 0.$$

Hence

$$I_p = 4s^p \sum_{k=0}^{p-1} Q^{(4k)}(1) a^{4(p-k-1)} = 4 \sum_{k=0}^{p-1} Q^{(4k)}(1) s^{k+1} (-r)^{p-k-1}. \tag{3}$$

Because

$$(x^4 - 1) = 4(x-1) + 6(x-1)^2 + 4(x-1)^3 + (x-1)^4,$$

we can write

$$Q(x) = \frac{1}{(p-1)!} \left(4^{p-1}(x-1)^{p-1} + b_1(x-1)^p + \cdots + b_{3p-3}(x-1)^{4p-4}\right),$$

where $b_1, \ldots, b_{3p-3}$ are integers. Thus

$$Q(1) = Q^{(1)}(1) = \cdots = Q^{(p-2)}(1) = 0, \quad Q^{(p-1)}(1) = 4^{p-1}, \tag{4}$$

and each of the quantities $Q^{(k)}(1)(k = p, p+1, \ldots)$ is divisible by p. Because $Q^{(p-1)}(1)$ is not divisible by p, it follows from (3) and (4) that I_p is not divisible by p, and therefore $|I_p| \geqslant 1$. But for large p this contradicts the inequality

$$|I_p| < \pi^{4p} s^p \frac{1}{(p-1)!} 4 e^{\pi},$$

which follows from (1).

In conclusion, maybe it will be possible to prove similarly the transcendency of π, but the author doesn't know how to do so.

Reference

1. Robert Breusch, A proof of the irrationality of π, this MONTHLY, 61 (1954) 631–632.

ON THE CONTINUITY OF VAN DER WAERDEN'S FUNCTION IN THE HÖLDER SENSE

ABDULLAH SHIDFAR
Department of Mathematics, Iran University of Science and Technology, Narmak, Tehran 16, Iran

KAZEM SABETFAKHRI
Department of Electrical Engineering, Iran University of Science and Technology, Narmak, Tehran 16, Iran

Van der Waerden's function is $f(x) = \sum_{k=0}^{\infty} a_k(x)$, where $a_0(x)$ is the distance from x to the nearest integer and $a_k(x) = 2^{-k} a_0(2^k x)$. It has been proved, in two recent notes, that this continuous function has no two-sided derivative at any point [**1**], nor does it have a one-sided derivative anywhere [**2**]. For $0 < s \leqslant 1$, recall that a function g is called Hölder continuous of class s if there is a constant M_s such that

$$|g(x+t) - g(x)| \leqslant M_s |t|^s$$

for all real numbers x and t.

Although the van der Waerden function f is nowhere differentiable, we will show here that it is Hölder continuous of class s for each s between 0 and 1. This is the best possible result in the sense that if the result held for $s = 1$, then f would be absolutely continuous and hence differentiable almost everywhere.

We first consider the case $0 < |t| \leqslant 1$. In this case, we can always find an integer n, with $n \geqslant 0$, such that

$$2^{-n-1} < |t| \leqslant 2^{-n}. \tag{1}$$

It can easily be shown that

$$|a_k(x+t) - a_k(x)| \leqslant |t| \tag{2}$$

for all real numbers x and t. From (1) and (2), we obtain

$$\sum_{k=0}^{n-1} |a_k(x+t) - a_k(x)| \leqslant n|t| \leqslant n 2^{-n(1-s)} |t|^s. \tag{3}$$

Moreover, since

$$0 \leqslant a_k(x) \leqslant 2^{-k-1}, \tag{4}$$

we have

$$|a_k(x+t) - a_k(x)| \leqslant 2^{-k-1}. \tag{5}$$

It follows from (1) and (5) that for any $|t| \leqslant 1$,

$$\sum_{k=n}^{\infty} |a_k(x+t) - a_k(x)| \leqslant 2^{-n} \leqslant 2|t| \leqslant 2|t|^s. \tag{6}$$

Summing inequalities (3) and (6), we obtain

$$\sum_{k=0}^{\infty} |a_k(x+t) - a_k(x)| \leqslant \{2 + n2^{-n(1-s)}\}|t|^s. \tag{7}$$

It is clear that there is a constant M_s such that

$$2 + n2^{-n(1-s)} \leqslant M_s \tag{8}$$

for every integer $n \geqslant 0$. Now (7) and (8) imply that

$$|f(x+t) - f(x)| \leqslant M_s|t|^s$$

whenever $0 < |t| \leqslant 1$.

Now we consider the case $|t| > 1$. It follows from (4) that

$$0 \leqslant f(x) \leqslant 1,$$

and for any $|t| > 1$, this yields

$$|f(x+t) - f(x)| \leqslant 1 \leqslant |t|^s$$

for any two real numbers x and t. Hence, choosing the same M_s as above ensures that

$$|f(x+t) - f(x)| \leqslant M_s|t|^s$$

for any t, and the proof is complete.

References

1. Patrick Billingsley, Van der Waerden's continuous nowhere differentiable function, this MONTHLY, 89 (1982) 691.

2. F. S. Cater, On Van der Waerden's nowhere differentiable function, this MONTHLY, 91 (1984) 307–308.

A NOTE ON EASY PROOFS OF STIRLING'S THEOREM*

COLIN R. BLYTH

Department of Mathematics, Queen's University, Kingston, Ontario, Canada K7L 3N6

PRAMOD K. PATHAK

Department of Mathematics, University of New Mexico, Albuquerque, NM 87131

Introduction. For Stirling's Theorem

$$\lim_{n\to\infty} \frac{\sqrt{2\pi}\, e^{-n} n^{n-.5}}{\Gamma(n)} = 1 \tag{1}$$

or, equivalently, multiplying numerator and denominator each by n,

$$\lim_{n\to\infty} \frac{\sqrt{2\pi}\, e^{-n} n^{n+.5}}{n!} = 1.$$

*This work was supported by NSERC, the National Sciences and Engineering Research Council of Canada, and by National Science Foundation Grant INT-8020450.

Khan [2] and Wong [5] give easy proofs valid for integers n only, by applying the Central Limit Theorem to Gamma and Poisson random variables respectively.

In Sections 1 and 2 these proofs are shortened; also they are made more elementary and Khan's extended to non-integers n by noticing that, not the Central Limit Theorem, but only the limit theorem for moment generating functions (or characteristic functions), is needed.

Section 3 gives an even shorter and more elementary proof by applying the inversion theorem for characteristic functions to a Gamma random variable. This method also has the advantage that it can be used (but the only proofs we have are too long and complicated to be satisfactory) to prove the extended Stirling Theorem

$$\frac{\Gamma(n)}{\sqrt{2\pi}\, n^{n-.5}e^{-n}} = 1 + \frac{1}{12n} + \frac{1}{288n^2} - \frac{139}{51840n^3} + O\left(\frac{1}{n^4}\right). \tag{2}$$

1. Khan's Proof. Because it uses the Central Limit Theorem, Khan's [2] proof is valid for integers n only. By using instead the limit theorem for moment generating functions, his proof can be made valid for non-integers as well. This proof begins by calculating the average deviation of a Gamma(n) random variable:

LEMMA 1. *If the random variable X has Gamma(n) probability density, with $n > 0$ not necessarily an integer, given by*

$$\frac{1}{\Gamma(n)} x^{n-1}e^{-x} \quad \text{for } x > 0,$$

then for $Z = (X-n)/\sqrt{n}$ we have

$$E|Z| = \frac{2e^{-n}n^{n-.5}}{\Gamma(n)}. \tag{3}$$

Proof. (A shorter proof of (3) than Khan's).

$$\begin{aligned} E|X-n| &= \frac{1}{\Gamma(n)}\int_0^\infty |x-n|x^{n-1}e^{-x}\,dx \\ &= -\frac{1}{\Gamma(n)}\int_0^n (x-n)x^{n-1}e^{-x}\,dx + \frac{1}{\Gamma(n)}\int_n^\infty (x-n)x^{n-1}e^{-x}\,dx \\ &= -\frac{-2}{\Gamma(n)}\int_0^n (x-n)x^{n-1}e^{-x}\,dx + \frac{1}{\Gamma(n)}\int_0^\infty (x-n)x^{n-1}e^{-x}\,dx. \end{aligned}$$

This last integral is $E(X-n) = EX - n = 0$, giving

$$E|X-n| = -\frac{2}{\Gamma(n)}\int_0^n x^n e^{-x}\,dx + \frac{2}{\Gamma(n)}\int_0^n nx^{n-1}e^{-x}\,dx.$$

Integrating the first of these integrals by parts gives

$$E|X-n| = \frac{2}{\Gamma(n)}e^{-n}n^n$$

and division by $\sqrt{n}$ gives (3), completing the proof of the lemma.

Now, to prove Stirling's Theorem, consider the moment generating function of Z which is given, for $t/\sqrt{n} < 1$, by

$$Ee^{t(X-n)/\sqrt{n}} = e^{-t\sqrt{n}}\frac{1}{\Gamma(n)}\int_0^\infty e^{tx/\sqrt{n}}e^{-x}x^{n-1}\,dx$$

$$= e^{-t\sqrt{n}}(1 - t/\sqrt{n})^{-n} = e^{-t\sqrt{n} - n\log(1 - t/\sqrt{n})}$$

$$= \exp\left\{\frac{t^2}{2} + O(1/\sqrt{n})\right\}.$$

As $n \to \infty$, this converges to $\exp(t^2/2)$ for every t, giving the moment generating function of the Standard Normal distribution, showing that the distribution function of Z converges to the Standard Normal. Moreover $EZ^2 = 1$ for every n, so the moment convergence theorem (Loève [3], p. 184) shows that $E|Z|$ must converge to the first absolute moment $\sqrt{(2/\pi)}$ of the Standard Normal. From (3),

$$\lim_{n\to\infty} \frac{2e^{-n}n^{n-.5}}{\Gamma(n)} = \sqrt{\frac{2}{\pi}}.$$

Division by $\sqrt{(2/\pi)}$ now gives Stirling's theorem (1).

NOTE. Having proved Stirling's theorem over integer and non-integer values n, we can now easily conclude for fixed a that

$$\lim_{n\to\infty} \frac{\Gamma(n+a)}{n^a\Gamma(n)} = 1, \tag{4}$$

instead of proving (4) and using it to derive Stirling's Theorem as Titchmarsh [4], page 58, does.

2. Wong's Proof. Wong's [5] proof of Stirling's theorem is valid for integers n only and can not be extended to non-integers. His proof can be made slightly more elementary by noticing that, not the Central Limit Theorem, but only the limit theorem for moment generating functions is needed; it can be made easier by using the more familiar $E|X-n|$ instead of $E(X-n)^-$ which eliminates the need to appeal to the Tucker theorem. This proof begins by calculating the average deviation of a Poisson (n) random variable.

LEMMA 2. *If the random variable X has Poisson (n) distribution with $n > 0$ an integer, then for $Z = (X-n)/\sqrt{n}$ we have*

$$E|Z| = \frac{2e^{-n}n^{n+.5}}{n!}. \tag{5}$$

Proof. This proof of (5) is a discrete version of the proof of (3):

$$E|X-n| = \sum_{x=0}^{\infty} |x-n|e^{-n}\frac{n^x}{x!}$$

$$= -e^{-n}\sum_{x=0}^{n-1}(x-n)\frac{n^x}{x!} + e^{-n}\sum_{x=n}^{\infty}(x-n)\frac{n^x}{x!}$$

$$= -2e^{-n}\sum_{x=0}^{n-1}(x-n)\frac{n^x}{x!} + e^{-n}\sum_{x=0}^{\infty}(x-n)\frac{n^x}{x!}.$$

This last sum is $E(X-n) = EX - n = 0$. The sum preceding it telescopes to give

$$E|X-n| = \frac{2e^{-n}n^n}{(n-1)!}$$

and division by $\sqrt{n}$ gives (5), completing the proof of the lemma.

Now, to prove Stirling's Theorem, consider the moment generating function of Z, which is given by

$$Ee^{t(X-n)/\sqrt{n}} = e^{-t\sqrt{n}} \sum_{x=0}^{\infty} e^{tx/\sqrt{n}} n^x \frac{e^{-n}}{x!}$$

$$= e^{-t/\sqrt{n}} \exp\{ n(e^{t/\sqrt{n}} - 1)\} = \exp\left\{ \frac{t^2}{2} + O(1/\sqrt{n})\right\},$$

and Stirling's theorem (1) now follows in exactly the same way as in Section 1.

NOTE. If this proof is tried with n not an integer, it gives the same result but with n replaced by $[n]$, the greatest integer in n.

3. A New Proof. This proof is shorter than the preceding two, is slightly more elementary, and is valid for arbitrary n:

If Z has continuous probability density function $f(\ )$ and integrable characteristic function $\phi(\ )$, then the inversion theorem for characteristic functions, Chung [**1**], p. 143, gives

$$f(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi(t)\, dt. \tag{6}$$

For the Z of Section 1 we have

$$f(0) = \frac{n^{n-.5} e^{-n}}{\Gamma(n)}$$

and, just as done in Section 1 for the moment generating function,

$$\phi(t) = e^{-it\sqrt{n}}(1 - it/\sqrt{n})^{-n} = e^{-t^2/2 + O(1/\sqrt{n})}.$$

Moreover, $\phi(t)$ is easily seen to be dominated by $1/(1 + t^2/2)$ for $n \geq 2$, so the dominated convergence theorem gives

$$\lim_{n\to\infty} \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi(t)\, dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-t^2/2}\, dt = 1/\sqrt{2\pi}.$$

Taking limits of the two sides of (6) for this Z therefore gives

$$\lim_{n\to\infty} \frac{n^{n-.5} e^{-n}}{\Gamma(n)} = 1/\sqrt{2\pi},$$

and multiplication by $\sqrt{2\pi}$ now gives Stirling's Theorem (1).

NOTE. It would be nice to have a proof of the extended Stirling Theorem (2) that is comparable in ease to these proofs of (1). The proof of this section, by using an expansion of $\phi(t)$ in (6), can be used to get a proof of (2), but the proof as we have it is too long and complicated to be satisfactory.

References

1. K. L. Chung, A Course in Probability Theory, Harcourt, Brace & World, New York, 1968.
2. R. A. Khan, A probabilistic proof of Stirling's formula, this MONTHLY, 81 (1974) 366–369.
3. M. Loève, Probability Theory, 3rd ed., Van Nostrand, Princeton, 1963.
4. E. C. Titchmarsh, The Theory of Functions, 2nd ed., Oxford University Press, London, 1939.
5. C. S. Wong, A note on the Central Limit Theorem, this MONTHLY, 84 (1977) 472.

ANSWER TO PHOTO ON PAGE 373

Richard Courant.

Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T: Textbook	P: Professional Reading	1-4: Semesters
C: Computer Software	L: Undergraduate Library	** : Special Emphasis
S: Supplementary Reading	13: Grade Level	?? : Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the Monthly.

General, P, L*.** Proceedings of the International Congress of Mathematicians, August 16-24, 1983, Warszawa. Ed: Czeslaw Olech, Zbigniew Ciesielski. Elsevier Science, 1984, $95 set [ISBN: 0-444-86661-2]. Volume 1, lxii + 848 pp; Volume 2, x + 881 pp. Reports on the work of the 1982 Fields medallists (Alain Connes, William Thurston, Shing-Tung Yau) and the Nevanlinna Prize awardee (Robert Tarjan); 12 of 16 invited one-hour plenary addresses; and 123 of 129 45-minute addresses in the 19 sections of the Congress. A timely publication just in advance of the 1986 ICM in Berkeley. LAS

Education, P, L. Learning and Mathematics Games. George W. Bright, John G. Harvey, Margariete Montague Wheeler. JRME Mono., No. 1. NCTM, 1985, ix + 188 pp, $7.50 (P). Presents research results on "cognitive effects of mathematical instructional games." Emphasis is three-fold: 1) summarizes existing research, 2) details of authors' research, based on instructional and taxonomic levels, and 3) conclusions drawn based on research results. Comprehensive presentation. RD

Education, L. Teacher-made Aids for Elementary School Mathematics, Volume 2: Readings from the Arithmetic Teacher. Ed: Carole J. Reesink. NCTM, 1985, v + 185 pp, $8 (P). [ISBN: 0-87353-225-2] A second collection of reprints from the Arithmetic Teacher, organized by skill areas. Most aids are for computational skills; few for probability or problem solving; none at all for estimation. LAS

History, P, L*. From One to Zero: A Universal History of Numbers. Georges Ifrah. Transl: Lowell Bair. Viking Pr, 1985, xvi + 503 pp, $35. [ISBN: 0-670-37395-8] A fascinating, intensely detailed exploration of the pre-historic origin of numbers: concrete counting, invention of numerals, alphabetic numerals, place value, ..., climaxing with the origin of the Hindu-Arabic numerals. Includes richly-illustrated samples of early numeration systems from diverse cultures--Sumerian, Hittite, Aztec, Chinese, Mayan, and many more. Amazingly, there is no index! Translated from the 1981 French original. LAS

History, S*(13-16), L*. A History of Mathematics, Fourth Edition. Florian Cajori. Chelsea, 1985, viii + 524 pp, $22.50. [ISBN: 0-8284-1303-X] The chapter on Babylonian mathematics has been rewritten to reflect recent research. A few minor improvements have been made in the text and in the Editor's Notes. JK

Foundations, T(16-18: 1, 2), S, P, L. Foundations of Constructive Mathematics: Metamathematical Studies. Michael J. Beeson. Ergebnisse der Math., 3, B. 6. Springer-Verlag, 1985, xxiii + 466 pp, $49. [ISBN: 0-387-12173-0] Organized into four parts: practice and philosophy of constructive mathematics, formal systems of the seventies, metamathematical studies, and metaphilosophical studies. Although the initial portions are accessible to anyone with theoretical analysis and advanced algebra, the later material presupposes a significant background in foundations. The engaging text is punctuated by numerous interesting exercises. MU

Graph Theory, P. Random Graphs. Béla Bollobás. Academic Pr, 1985, xvi + 447 pp, $58.50; $29.95 (P). [ISBN: 0-12-111755-3; 0-12-111756-1] A tour de force on probabilistic methods in graph theory, especially extremal questions. SS

Graph Theory, T, S, P, L. Lecture Notes in Economics and Mathematical Systems-250: Preference Modelling. Marc Roubens, Philippe Vincke. Springer-Verlag, 1985, viii + 94 pp, $12.30 (P). [ISBN: 0-387-15685-2] Assuming virtually no mathematical prerequisites, the authors use graphs, numbers, and tableau (and their relationships) to discuss "preference structures." This includes the usual structures (e.g., tournaments, total orders, partial orders), two new structures (partial interval order, partial semi-order), and a "valued" preference structure which appears in probabilistic contexts. LCL

Graph Theory, T*(15-17: 1), S, P, L. Algorithmic Graph Theory. Alan Gibbons. Cambridge U Pr, 1985, xii + 259 pp, $47.50; $17.95 (P). [ISBN: 0-521-24659-8; 0-521-28881-9] An introductory text,

intended primarily for computer scientists but of interest to mathematicians as well, which features a broad range of graph-theoretic algorithms: spanning trees, planarity, network flows, matching, touring, coloring. Includes discussion of complexity. LCL

Graph Theory, T(17-18: 1), S, P. Graphical Evolution: An Introduction to the Theory of Random Graphs. Edgar M. Palmer. Ser. in Disc. Math. Wiley, 1985, xvii + 177 pp, $34.95. [ISBN: 0-471-81577-2] Truly an introduction to random graphs, including relevant counting techniques, probabilistic models and threshold functions (used to study larger order graphs as the size increases). In only 115 pages, readers are led through fascinating mathematics to fascinating unsolved problems. Many exercises. SS

Combinatorics, S(17), P. Surveys in Combinatorics 1985. Ed: Ian Anderson. London Math. Soc. Lect. Note Ser., V. 103. Cambridge U Pr, 1985, vi + 173 pp, $18.95 (P). [ISBN: 0-521-31524-7] This volume contains eight of the nine invited lectures given at the Tenth British Combinatorial Conference held at the University of Glasgow in July, 1985. It provides a broad survey of many areas of contemporary research interest. CEC

Combinatorics, T(17-18: 1), S, P. Design Theory. D.R. Hughes, F.C. Piper. Cambridge U Pr, 1985, viii + 240 pp, $39.50. [ISBN: 0-521-25754-9] Designs, which are actually combinatorial structures, were originally conceived in statistics, but are now seen to have intimate connections with finite geometry, coding theory, graph theory and group theory. This introduction covers the classical theory and ties it to various applications. SS

Discrete Mathematics, S(17), P. Discrete Geometry and Convexity. Ed: Jacob E. Goodman, et al. Annals of NY Acad. of Sci., V. 440. NY Academy of Sciences, 1985, xii + 392 pp, $90 (P). [ISBN: 0-89766-275-X] An outgrowth of a meeting held in April 1982 at the New York Academy of Science and the Courant Institute. Expanded versions of nine of these talks appear here along with some additional papers which were solicited to demonstrate the scope of recent work in convexity and discrete geometry. CEC

Number Theory, T(17: 1), S, P, L. Ultrametric Calculus: An Introduction to p-adic Analysis. W.H. Schikhof. Cambridge U Pr, 1984, xi + 306 pp, $39.50. [ISBN: 0-521-24234-7] This text requires an elementary understanding of algebra and analysis. Starting with these basics the author presents a complete development of p-adic analysis which parallels a standard course in analysis and displays a variety of applications of the subject. Includes a large number of exercises. CEC

Number Theory, P. Séminaire de Théorie des Nombres, Paris 1983-84: Séminaire Delange-Pisot-Poitou. Ed: Catherine Goldstein. Prog. in Math., V. 59. Birkhauser Boston, 1985, viii + 278 pp, $32.95. [ISBN: 0-8176-3315-4] Seventeen papers on diverse topics in number theory, expositing work published or to appear elsewhere. BC

Number Theory, S(17), P. Analytic Methods in the Analysis and Design of Number-Theoretic Algorithms. Eric Bach. ACM Dist. Dissertation 1984. MIT Pr, 1985, 48 pp, $15. [ISBN: 0-262-02219-2] This series recognizes doctoral dissertations of such high quality that they should be published. The author presents a derivation of the minimum effort required to test if an integer is prime and then attacks the problem of generating a large random number with known factors. Both results are useful in cryptography. CEC

Group Theory, P. Proceedings of the Rutgers Group Theory Year, 1983-1984. Ed: Michael Aschbacher, et al. Cambridge U Pr, 1984, xii + 415 pp, $39.50. [ISBN: 0-521-26493-6] Forty-four papers concerning revision of the classification of finite simple groups; development of the properties of simple groups; applications of the classification to finite group theory, number theory and geometry; chamber systems and amalgams; computational algorithms for groups. DFA

Algebra, S(16-18), L. Algebra through Practice: A Collection of Problems in Algebra with Solutions, Books 4, 5 & 6. T.S. Blyth, E.F. Robertson. Cambridge U Pr, 1985, x + 100 pp, $39.50. [ISBN: 0-521-25301-2] Second in a series, this volume is a collection of problems dealing with linear algebra, groups, rings, fields, and modules. The problems tend to be of medium difficulty with very few routine or computational. Solutions are included, as well as three-hour test papers for each chapter. Background references are listed, mostly to standard texts. JS

Algebra, P. Lecture Notes in Mathematics-1149: Universal Algebra and Lattice Theory. Ed: Stephen D. Comer. Springer-Verlag, 1985, vi + 282 pp, $17.60 (P). [ISBN: 0-387-15691-7] Proceedings of a conference held at the Citadel in Charleston, South Carolina in 1984, with papers containing new results and representing new directions of research. LCL

Algebra, T*(1), S*, P*, L*. Contemporary Abstract Algebra. Joseph A. Gallian. DC Heath, 1986, xviii + 468 pp, $29.95. [ISBN: 0-669-09325-4] Everything that one would expect in a beginning course in abstract algebra is here, and yet, so much more. Lots of exercises, including computer exercises, applications, twenty-four biographical sketches, references and suggested reading, all sorts of quotes, and a sense of humor. Looking at this book makes you look forward to teaching your next algebra class. CEC

Calculus, T(13: 2). Calculus and Analytic Geometry, Second Edition. C.H. Edwards, Jr., David E. Penney. Prentice-Hall, 1986, xiii + 1086 pp, $39.95. [ISBN: 0-13-111675-4] This edition includes additional drill exercises, more elementary examples, increased detail in existing examples, subdivision and rearrangement of several sections, introduction of optional programming notes and com-

puter graphics. All of this adds 200 pages. (First Edition, TR, November 1982; Extended Review, October 1983.) JNC

Calculus, C?(12-13). EPIC: Exploration Programs in Calculus. IBM PC. James W. Burgmeier, Larry L. Kost. Prentice Hall, 1985, $34.95. A "turnkey" menu-driven package of graphics routines for simple functions, simple limits, special cases of Riemann sums, polar and parametric equations. Function entry is somewhat awkward, probably caused by the limitations of what appears to be compiled BASIC. Within rather narrow limits, the graphics is neat and quickly done. However, the habit the program has of counting backwards in a corner of the screen and then flashing the completed graph on the screen seems to emphasize the magic of the computer rather than the computation of a graph which is seen developing on the screen. Supporting document is clear, complete, and well-illustrated. Requires only 64K memory, one disk drive, and the graphics card; uses its own version of the PC-DOS operating system and will not work execpt in its own environment. JAS

Calculus, S*(13). Quick Calculus, Second Edition: A Self-Teaching Guide. Daniel Kleppner, Norman Ramsey. Wiley, 1985, x + 262 pp, $10.95 (P). [ISBN: 0-471-82722-3] For self-instruction. Emphasizes technique and application without rigor. Second Edition incorporates use of calculators, more examples and more applications. First Edition sold over 250,000 copies. Might be worth a look. "The idea for it grew out of the problem of teaching college freshmen enough calculus so that they could start physics without waiting for a calculus course in college." JK

Calculus, T*(13-14: 3). Calculus with Analytic Geometry, Third Edition. Roland E. Larson, Robert P. Hostetler. DC Heath, 1986, xxi + 1187 pp, $39.95. [ISBN: 0-669-09568-0] Incorporates many changes suggested by users of the earlier editions (First Edition, TR, June 1979; Second Edition, TR, October 1982). Coverage of trigonometric functions has been moved into the first six chapters. There is an alternate edition which preserves the organization of the earlier editions. JK

Real Analysis, T, L. Metric Spaces. Victor Bryant. Cambridge U Pr, 1985, vi + 105 pp, $29.95; $9.95 (P). [ISBN: 0-521-26857-5; 0-521-31897-1] Beginning with an application--contractions for finding solutions to equations--the author motivates the need for distance, function spaces, closed and open sets. An intriguing approach and a very readable text. TAV

Partial Differential Equations, P. Théorie de la Deuxième Microlocalisation dans le Domaine Complexe. Yves Laurent. Progress in Math., V. 53. Birkhauser Boston, 1985, xvi + 311 pp, $34.95. [ISBN: 0-8176-3287-5] Studies the sheaf of rings of "2-microdifferential operators," defined on the cotangent bundle of a Lagrangian submanifold of a complex manifold. The relation between such objects and microdifferential operators is analogous to that between microdifferential and differential operators. Applications to complex analysis appear in the last chapter. PZ

Partial Differential Equations, P. Lecture Notes in Mathematics-1127: Numerical Methods in Fluid Dynamics. Ed: F. Brezzi. Springer-Verlag, 1985, vii + 333 pp, $20.50 (P). [ISBN: 0-387-15225-3] Lectures given at Como, Italy in July 1983. R. Glowinski and J. Periaux on finite element, least squares and domains decomposition methods; G. Gottlieb and Eli Turkel on spectral methods; A. Jameson on transonic flow calculations; P.A. Raviart on particle methods. DFA

Partial Differential Equations, P. Schauder's Estimates and Boundary Value Problems for Quasilinear Partial Differential Equations. Manfred König. Pr U Montreal, 1985, 141 pp, $18 (P). [ISBN: 2-7606-0694-5] Lecture notes from a 1983 NATO science seminar at the University of Montreal. The first chapter studies existence, uniqueness, and regularity of solutions to boundary value problems for uniformly elliptic differential operators, mainly by reducing such problems to classical Dirichlet form. The second chapter concerns linear and nonlinear radiation problems. PZ

Partial Differential Equations, P. Lecture Notes in Mathematics-1151: Ordinary and Partial Differential Equations. Ed: B.D. Sleeman, R.J. Jarvis. Springer-Verlag, 1985, xiv + 357 pp, $23.50 (P). [ISBN: 0-387-15694-1] Proceedings of the eighth conference on ordinary and partial differential equations held at the University of Dundee, Scotland, June 25-29, 1984. Contains twenty-five papers addressing the theory of nonlinear differential equations and their relevance to biological phenomena and nonlinear wave propagation. AM

Numerical Analysis, S(18), P. Numerical Solutions of the N-body Problem. Andrzej Marciniak. Math. & Its Applic. D Reidel, 1985, xi + 242 pp, $39.50. [ISBN: 90-277-2058-4] Reviews conventional numerical methods for solving initial value problems; for the N-body problem, these methods violate conservation laws. Develops special "discrete mechanics" methods which obey conservation. Many algorithms. Caution: the symbol "l" is used simultaneously as a variable and as the number "one." BC

Numerical Analysis, T?(18), P. Multi-Grid Methods and Applications. Wolfgang Hackbusch. Ser. in Comput. Math., V. 4. Springer-Verlag, 1985, xiv + 377 pp, $59. [ISBN: 0-387-12761-5] Introduction to a class of fast-converging algorithms for elliptic boundary problems. Algorithmic details, rigorous convergence analysis, special applications to fluid dynamics, perturbation, eigenvalue problems, integral equations. RM

Numerical Analysis, T(17-18: 1, 2), S, P*, L. Numerical Solution of Elliptic Problems. Garrett Birkhoff, Robert E. Lynch. Stud. in Appl. Math. SIAM, 1984, xi + 319 pp, $31.50. [ISBN: 0-89871-197-5] The use and analysis of difference, iterative, integral and finite elements methods of solving elliptic boundary value problems computationally, especially in two-dimensions. Includes background on physical origins, classical analysis, linear algebra, and the package ELLPACK. RWN

Numerical Analysis, T(18), P. Analysis of Approximation Methods for Differential and Integral Equations. H.-J. Reinhardt. Appl. Math. Sci., V. 57. Springer-Verlag, 1985, xi + 398 pp, $45 (P). [ISBN: 0-387-96214-X] The "consumer" of approximation methods may prefer to take convergence for granted, but somebody has to worry about how good the answers are. Reinhardt does, developing a general theory based on "discrete convergence" and applying it to various problems. Intended for advanced students and researchers. Thorough proofs and examples, no exercises. BC

Numerical Analysis, P. Progress and Supercomputing in Computational Fluid Dynamics. Ed: Earll M. Murman, Saul S. Abarbanel. Prog. in Sci. Comp., V. 6. Birkhauser Boston, 1985, ix + 403 pp, $44.95. [ISBN: 0-8176-3321-9] Proceedings of a joint U.S.-Israel workshop held in Jerusalem during December 1984, introduced by an overview paper written by the editors that summarizes the major themes that emerged from the workshop. A good state-of-the-art survey of one of the most significant and common applications of supercomputers. LAS

Functional Analysis, S(18), P. Unbounded Linear Operators: Theory and Applications. Seymour Goldberg. Dover, 1985, viii + 199 pp, $6 (P). [ISBN: 0-486-64830-3] An unabridged, corrected edition of the original work published by McGraw-Hill in 1966 (TR, January 1967; Extended Review, August-September 1968). JS

Functional Analysis, P. Lecture Notes in Mathematics-1131: Geometry and Nonlinear Analysis in Banach Spaces. Kondagunta Sundaresan, Srinivasa Swaminathan. Springer-Verlag, 1985, 116 pp, $9.80 (P). [ISBN: 0-387-15237-7] Classical theory and recent developments in the study of non-linear problems (e.g., smoothness of norms, partitions of unity, approximation theory) in Banach spaces. The degree to which classical theorems (e.g., Stone-Weierstrass, Bernstein) carry over to the infinite-dimensional case is an important theme. An appendix treats manifolds modelled on Banach spaces. PZ

Functional Analysis, T(18: 1), S, P. Positive Operators. Charalambos D. Aliprantis, Owen Burkinshaw. Pure & Appl. Math., V. 119. Academic Pr, 1985, xvi + 367 pp, $59. [ISBN: 0-12-050260-7] Aimed toward extending (with minimal duplication) earlier books by Schaefer and Zaanen, topics include order structure, components, homomorphisms, Banach lattices, and compactness. Exercises, references, index. JS

Functional Analysis, S*(18), P.** Metric Linear Spaces. Stefan Rolewicz. Math. & Its Applic. D Reidel, 1985, xii + 459 pp, $79. [ISBN: 90-277-1480-0] An encyclopedic, definitive treatment of non-locally convex spaces. Even with some simple problems with the translation to English, this is a wonderfully complete treatment. Huge bibliography, no exercises. (1972 PWN Edition, TR, December 1973.) TAV

Analysis, T(18), S, P. Transmutation Theory and Applications. Robert Carroll. Math. Stud., V. 117. Elsevier Science, 1985, xii + 351 pp, $35 (P). [ISBN: 0-444-87805-X] This monograph begins with an introduction to distributions, Fourier analysis and spectral theory, and develops the transmutation machinery in detail. It contains new and recent material relative to characterization of transmutations by minimization, by Cauchy and Goursat problems, by spectral pairings, by connections of generalized eigenfunctions, by domains, etc. MU

Analysis, P. Méthodes Topologiques en Analyse Non Linéaire. Andrzej Granas. Pr U Montreal, 1985, 235 pp, $22 (P). [ISBN: 2-7606-0712-7] Proceedings of a June 1983 NATO advanced study institute held at the University of Montreal. Eight papers on topological analysis, including a lengthy survey of generalizations of the Borsuk-Ulam antipodal theorem. LAS

Analysis, P. Trends in the Theory and Practice of Non-Linear Analysis. Ed: V. Lakshmikantham. Math. Stud., V. 110. Elsevier Science, 1985, x + 491 pp, $55 (P). [ISBN: 0-444-87704-5] About 60 short research and survey papers treating theory, applications, and computational aspects. Proceedings of the Sixth International Conference on Trends in the Theory and Practice of Non-linear Analysis held at Arlington, Texas in June 1984. PZ

Analysis, S*(15-16), L*. Integral Equations. F.G. Tricomi. Dover, 1985, viii + 238 pp, $6 (P). [ISBN: 0-486-64828-1] Unabridged and unaltered republication of Interscience's 1957 edition. Once "the" introductory text on integral equations. Not too rigorous. Not too abstract. Accessible to those needing the tool of integral equations. The usual introductory and advanced calculus courses are sufficient for prospective readers. JK

Algebraic Geometry, P. Lecture Notes in Mathematics-1137: Surfaces fibrées en courbes de genre deux. Xiao Gang. Springer-Verlag, 1985, ix + 103 pp, $9.80 (P). [ISBN: 0-387-15662-3]

Algebraic Geometry, T(17: 2), S, P. Introduction to Algebraic Geometry. Ed: J.G. Semple, L. Roth. Oxford U Pr, 1985, xviii + 454 pp, $21.95 (P). [ISBN: 0-19-853363-2] A re-issue, with index and corrections, of the well-known 1949 volume. Excellent classical background--results, techniques, and examples--for the modern practitioner of algebraic geometry. SG

Differential Geometry, S(17-18), P. Lecture Notes in Mathematics-1160: The Hamiltonian Hopf Bifurcation. Jan-Cees van der Meer. Springer-Verlag, 1985, vi + 115 pp, $9.80 (P). [ISBN: 0-387-16037-X] Given a vector field depending on a parameter, bifurcation theory refers to the study of qualitative changes in the flow of the vector field as the parameter crosses a critical value. Describes the bifurcation theory of Hamiltonian vector fields and methods for deriving normal forms for Hamiltonian systems around stable points. Contains good discussions of prerequisites from symplectic geometry and singularity theory. AM

Geometry, S(15-16), L. Geometrical Algebra. Paolo Bonasoni. Golden Hind Pr, 1985, 202 pp. [ISBN: 0-931267-01-3] Except for a still unpublished treatise on the division of circles, this is the only known work of the Renaissance mathematician, Bonasoni, circa 1575. Italian text with translation. Typical question is "to discover two straight lines in the proportion of BC to CD so that the squares of the lines themselves, together with the rectangle contained under these same lines, are equal to a rectilinear figure F." In a separate pocket-enclosed editorial the translator, Robert Schmidt, describes Bonasoni's method as "exactly reversing the algebraic argument" and cites Bonasoni's work as "the only example of an algebraic symbolism without an algebraic notation. JK

Geometry, S*(14), P, L. Groups: A Path to Geometry. R.P. Burn. Cambridge U Pr, 1985, xii + 242 pp, $49.50. [ISBN: 0-521-30037-1] Group theory is developed in a sequence of 800 problems concentrating on geometric transformations: chapters include introductory remarks, historical notes, suggestions for concurrent reading, answers and summaries of theorems and definitions. JNC

Geometry, T(16: 1), L.** Fundamental Concepts of Geometry. Bruce E. Meserve. Dover, 1983, ix + 321 pp, $7.50 (P). [ISBN: 0-486-63415-9] "An unabridged and slightly corrected republication of the second printing of the work originally published by Addison-Wesley...in 1955;" a classic text at a reasonable price. JNC

Algebraic Topology, S(18), P. Lecture Notes in Mathematics-1144: Knot Theory and Manifolds. Ed: D. Rolfsen. Springer-Verlag, 1985, 163 pp, $12 (P). [ISBN: 0-387-15680-1] These notes are based on eleven lectures given at the special session on knot theory and manifolds at the summer 1983 meeting of the Canadian Mathematical Society. CEC

Differential Topology, P. Symposium on Anomalies, Geometry, Topology. Ed: William A. Bardeen, Alan R. White. World Scientific, 1985, xviii + 558 pp, $65. [ISBN: 9971-978-69-5] Collection of 55 papers presented during a symposium sponsored by Argonne National Laboratory, Fermi National Accelerator Laboratory, and the University of Chicago, March 28-30, 1985. Presentors included M. Atiyah, E. Witten, and S.T. Yau. AM

Operations Research, T(15: 1), S, L. An Introduction to Linear Programming, Second Edition. G.R. Walsh. Wiley, 1985, ix + 240 pp, $22.95 (P). [ISBN: 0-471-90719-7] This edition adds a chapter on the ellipsoid method to the previous coverage of the simplex method, duality and the revised simplex method and applications to problems of transportation and assignment and relationship of linear programming to the theory of games; presumes linear algebra. JNC

Optimization, T(16-17: 2), P, L. Theory of Multiobjective Optimization. Yoshikazu Sawaragi, Hirotaka Nakayama, Tetsuzo Tanino. Math. in Sci. & Eng., V. 176. Academic Pr, 1985, xiii + 296 pp, $48. [ISBN: 0-12-620370-9] The authors attempt to provide a complete unified theory of multiobjective optimization. A few applications are presented, but the theory is the objective and it is well met. No exercises, substantial bibliography. TAV

Optimization, T(17-18: 1, 2), P*. The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization. Ed: E.L. Lawler, et al. Ser. in Disc. Math. Wiley, 1985, x + 465 pp, $64.95. [ISBN: 0-471-90413-9] A collection of twelve papers on various aspects of the traveling salesman problem each written by an expert. An excellent survey of the state-of-the-art. AO

Dynamical Systems, S(17-18), P. Chaos in Dynamic Systems. G.M. Zaslavsky, L.V. Kirenskii. Transl: V.I. Kisin. Harwood Academic, 1985, xix + 370 pp, $195. [ISBN: 3-7186-0225-3] Chaos, or stochasticity, refers to the random variation of the properties that describe a dynamical system, specifically random variations that arise due to instabilities characteristic of the dynamical system. An important issue is the extent to which this phenomenon is typical of general dynamic systems. This book discusses work done in the last 20 years and its applications. Includes discussions of KAM theory, billiards and other simple models, and generation of chaos in Hamiltonian systems. AM

Probability, T(17: 1), P. Stationary Sequences and Random Fields. Murray Rosenblatt. Birkhauser Boston, 1985, 258 pp, $34.95. [ISBN: 3-7643-3264-6] Presents material in time series analysis from both the finite parameter and spectral approach. Leads the reader from elementary concepts to open research questions in a smooth, readable way. Discusses asymptotic normality, structure functions, ARMA processes, and Monte Carlo methods. Extensive bibliography, no exercises. TAV

Statistics, T(13: 1). Statistics, A First Course, 4th Edition. John E. Freund, Richard Manning Smith. Prentice-Hall, 1986, xiv + 557 pp, $28.95. [ISBN: 0-13-845975-4] Some revision of several sections along with new material on stem-and-leaf plots, box-and-whisker diagrams, and tests with paired data in this new edition. Presupposes only high school algebra. (First Edition, TR, August-September 1970; Second Edition, May 1976; Third Edition, TR, March 1981.) FLW

Statistics, P. Statistics and Control of Stochastic Processes. Ed: N.V. Krylov, R. Sh. Lipster, A.A. Novikov. Steklov Seminar 1984. Optimization Software, 1985, xiii + 507 pp, $54. [ISBN: 0-911575-18-9] Proceedings of a 1984 seminar at the Steklov Institute of Mathematics in Moscow, dedicated to A.N. Shiryaev on his 50th birthday. LAS

Statistics, P. Lecture Notes in Statistics-27: Infinitely Divisible Statistical Experiments. Arnold Janssen, Hartmut Milbrodt, Helmut Strasser. Springer-Verlag, 1985, vi + 164 pp, $13.20 (P). [ISBN: 0-387-96055-4] Exposition of ideas presented in 1974 by L. LeCam on the asymptotic theory of statistics: "Roughly speaking, we deal with limits of products of independent experiments." LAS

Statistics, C(15-16). Introductory Simulation and Statistics Package. IBM PC. P.A.W. Lewis, Endel J. Orav, Luis Uribe. Wadsworth, 1984, $39.95. Compiled Fortran subroutines to analyse the output of user-written simulation programs. LAS

Computer Literacy, T(13-14: 1). Computers Today and Tomorrow: The Microcomputer Explosion. Tom Logsdon. Computer Science Pr, 1985, xiii + 361 pp, $26.95. [ISBN: 0-88175-026-3] A text for microcomputer literacy. Includes hardware and software evolution, societal aspects (electronic privacy intrusion, computer crime, robotics, CAI), and a look at the future. Chapters include exercises and student projects. Profusely illustrated with descriptive drawings, graphs, charts, and tables. RD

Computer Programming, T(13-14). Advanced BASIC, Step by Step. Vern McDermott, Diana Fisher. Computer Science Pr, 1984, xi + 315 pp, $29.96. [ISBN: 0-88175-011-5] Intended as an advanced text in BASIC programming. Major portion covers aspects of using sequential and random access files; also chapters on substrings, formatting, subroutines, and matrices. Lessons include statement of objectives, examples, explanations, and sample problems. Exercises are provided. RD

Computer Programming, T(13), S. The First Book of Macintosh Pascal. Paul A. Sand. Osborne McGraw-Hill, 1985, xvi + 411 pp, $17.95 (P). [ISBN: 0-07-881165-1] An introduction to common use of MacPascal, the interactive multi-windowed interpreted version of PASCAL that runs on the Macintosh computer. Covers selected parts of standard Pascal plus the special MacPascal extensions (notably graphics). Intended to supplement the detailed software reference manual. Skimpy table of contents and lack of summary tables in Appendices make it difficult to use the book as a reference. LAS

Computer Programming, T. Introduction to Structured Programming Using Basic. Coleman Barnett. Gorsuch Scarisbrick, 1984, xiv + 505 pp, $24 (P). [ISBN: 0-89787-402-1] Emphasizes concepts, techniques, algorithm development. BASIC is used to demonstrate techniques applicable to many languages. Problem solving is a key aspect of presentations. Includes annotated examples, exercises, and discussion questions. RD

Computer Programming, T(13: 1). Introduction to Computer Mathematics. Russell Merris. Computers & Math Ser. Computer Science Pr, 1985, ix + 284 pp, $27.95. [ISBN: 0-88175-083-2] A miscellanea of elementary topics from probability, statistics, algebra and geometry is presented and used as sources for BASIC computer programs (e.g., the last section programs the trapezoidal rule). JNC

Computer Programming, P, L. Mastering the Macintosh Toolbox. David B. Peatroy. Osborne McGraw-Hill, 1986, viii + 200 pp, $16.95 (P). [ISBN: 0-07-881203-8] Detailed discussion of the Macintosh's internal collection of programming routines and hardware (mouse and screen) software. Uses MacPascal as the standard programming environment. Contains programs for editing icons, controlling the mouse, deploying windows, and designing screen displays. LAS

Computer Programming, S(13-16), L. The C Library. Kris Jamsa. Osborne McGraw-Hill, 1985, x + 294 pp, $18.95 (P). [ISBN: 0-07-881110-4] For those familiar with or in the process of learning C. Presents a summary of C and a set of definitions to provide a standard C environment. Following this there are nine chapters explaining algorithms and providing C-functions using those algorithms for string manipulation, I/O, array manipulation, sorting, recursion, file manipulation and using the UNIX/MS-DOS pipe. The book provides a nice collection of examples that extend the usual C-library functions. The presentation is quite system-independent, although a few of the more subtle MS-DOS/UNIX inconsistencies may not be adequately covered up to allow everything presented here to work without a little fiddling with I/O, at least for some implementations of C. JAS

Computer Programming, T(13: 1). Structured Cobol: Pseudocode Edition. Gary B. Shelly, Thomas J. Cashman, Steven G. Forsythe. Anaheim Pub, 1985, xiv + 503 pp, $24.95 (P). [ISBN: 0-88236-127-9] Complete introduction to COBOL emphasizing structured design and structured programming techniques. Examples are typical COBOL applications. Extensive coverage of tables. Data editing and validation. Sequential disk file I/O. One appendix contains many sets of test data for the chapter programming assignments; another introduces flow charts. DFA

Software Systems, P. Control Flow and Data Flow: Concepts of Distributed Programming. Ed: Manfred Broy. NATO ASI Ser. F: Comp. & Sys. Sci., V. 14. Springer-Verlag, 1985, viii + 525 pp, $59. [ISBN: 0-387-13919-2] Proceedings of a NATO Advanced Study Institute held during August 1984 at Marktoberdorf, West Germany. 14 major lectures, four by E.W. Dijkstra, on models, hardware, design and verification of distributed multiprocessor computer systems. LAS

Software Systems. PICK for Users. Martin Taylor. Computer Science Pr, 1985, xiii + 183 pp, $17.95 (P). [ISBN: 0-632-01492-X] Practical introduction to PICK, a multi-user operating system oriented to commercial (business and administrative) uses, available on many systems, micros to mainframes. PICK's database feature and English-like query language ACCESS allow simple production of reports. Job control language (PROC) and PICK BASIC allow general processing, with emphasis on data rather than programming. RM

Computer Science, T?(14-16: 1, 2), S. Principles of Computer Design. Leonard R. Marino. Princ. of Comp. Sci. Ser. Computer Science Pr, 1985, xiii + 578 pp, $37.95. [ISBN: 0-88175-064-6] Description of computer organization and design principles at all levels: electronics, logic, register transfer, architecture, programming, from designer's point of view. Introduction to design practice, with role of multi-level hierarchy, models, principles emphasized. RM

Computer Science, P. Lecture Notes in Computer Science-202: Rewriting Techniques and Applications. Ed: Jean-Pierre Jouannaud. Springer-Verlag, 1985, vi + 440 pp, $22.80 (P). [ISBN: 0-387-15976-2] Proceedings of the first international conference on rewriting (replacement, reduction rules) in Dijon, France, 1985. Includes theoretical papers (e.g., on termination), applications to unification, proof theory, Thue systems, word problems, Petri nets. RM

Computer Science, P. Reliable Computer Systems: Collected Papers of the Newcastle Reliability Project. Ed: S.K. Shrivastava. Texts & Mono. in Comp. Sci. Springer-Verlag, 1985, xii + 580 pp, $39.50. [ISBN: 0-387-15256-3] An organized re-presentation of papers originally published prior to 1983 by researchers at the University of Newcastle-upon-Tyne on system reliability; exception handling; recovery; and concurrent, distributed, and multi-level systems. LAS

Computer Science, T*(15-17: 1, 2), L. Computer Organization: Hardware/Software, Second Edition. G.W. Gorsline. Prentice-Hall, 1986, xvi + 623 pp, $40.95. [ISBN: 0-13-165325-3] A sophisticated text which discusses considerably more than the architecture and assembly language of one or two machines. The revisions occur througout and reflect changes in thinking about processor architecture, system and network architecture, and the blurring of system size distinctions. Examples are drawn from a wide variety of processors thus giving a flavor of the transition from yesterday to tomorrow. (First Edition, TR, June-July 1980.) JAS

Computer Science, T(16-17: 1), P. Communicating Sequential Processes. C.A.R. Hoare. Ser. in Comput. Sci. Prentice-Hall, 1985, 256 pp, $37.95. [ISBN: 0-13-153271-5] Mathematical theory for describing a range of systems, from vending machines to complex operating systems. Based on notion of synchronized unbuffered communication, gives abstract framework for specification, design, implementation, validation of computer systems. Process, concurrency, nondeterminism, communication, shared resources formalized mathematically. Important work, whose principles are closely followed in the simple programming language Occam. RM

Computer Science, T(14-16: 1). Computer Organization and Assembly Language Programming for the PDP-11 and VAX-11. Wen C. Lin. Harper & Row, 1985, xi + 371 pp. [ISBN: 0-06-044061-9] A relatively machine-specific presentation of the standard topics with an emphasis on a PDP-11 with a macro assembler (assuming the RSX-11M operating system where needed). The VAX material appears briefly at the end. JAS

Computer Science, L. Dictionary of Computing. Ed: Valerie Illingworth, et al. Oxford U Pr, 1983, xi + 393 pp, $24.95. [ISBN: 0-19-853905-3] Concise, up-to-date dictionary of computer science terms; theoretical, applied, software, hardware, etc. RM

Computer Science, P. Lecture Notes in Computer Science-195: A Hierarchical Associative Processing System. Heinrich J. Stüttgen. Springer-Verlag, 1985, xii + 273 pp, $16.60 (P). [ISBN: 0-387-15652-6] Discussion of two-level associative (content addressible) memory hierarchy for general purpose computers, as an alternative to the Von Neumann paradigm. Fundamentals of associative memory, PASCAL/A--extension of Pascal to include tables (associative relational data structure) and generic associative instructions, memory hierarchy management employing staging, and architecture. RM

Applications, P. New Systems and Architectures for Automatic Speech Recognition and Synthesis. Ed: Renato De Mori, Ching Y. Suen. NATO ASI Ser. F: Comp. & Sys. Sci., V. 16. Springer-Verlag, 1985, xiii + 629 pp, $56. [ISBN: 0-387-15177-X] Proceedings of a NATO Advanced Studies Institute held in Bonas, Gers, France during July 1984: basic algorithms, system architecture, software system, and speech synthesis. A state-of-the-art presentation of automated speech recognition and synthesis. LAS

Applications (Artificial Intelligence), P. Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project. Ed: Bruce G. Buchanan, Edward H. Shortliffe. Ser. in Artif. Intell. Addison-Wesley, 1985, xx + 748 pp. [ISBN: 0-201-10172-6] Detailed retrospective view of MYCIN (rule-based expert consultant system for diagnosis and therapy of bacterial infections), its development and extensions (e.g., EMYCIN, ONCOCIN). Original papers rewritten and augmented, plus much new material. Discussion of explanation, tutoring, knowledge engineering, etc. RM

Applications (Artificial Intelligence), P. Readings in Medical Artificial Intelligence: The First Decade. Ed: William J. Clancey, Edward H. Shortliffe. Ser. in Artif. Intell. Addison-Wesley, 1984, xvi + 512 pp. [ISBN: 0-201-10854-2] Collection of important papers on artificial intelligence in medicine (up to 1981), each introduced with brief discussion of its place in the discipline. Includes chapters on artificial intelligence and its relevance to medicine, future directions, research efforts (MYCIN, NEOMYCIN, INTERNIST, PUFF, etc.). RM

Applications (Biology), P. Lecture Notes in Biomathematics-58: Branching Processes Applied to Cell Surface Aggregation Phenomena. Catherine A. Macken, Alan S. Perelson. Springer-Verlag, 1985, viii + 122 pp, $11.40 (P). [ISBN: 0-387-15656-9] A long-standing problem in theoretical cell biology is to predict the distribution of sizes of aggregates that form when monomer units (e.g., cells) join together to form polymers or aggregates. The method of "branching processes" brings a new approach to this study. The authors trace the history, describe the method, and apply it to special cases of the above problem. The techniques developed in these notes should be useful in other "aggregation" problems, as well as in other areas of biology. LCL

Applications (Communication Theory), T(16: 1), S, P, L. Basics of Communications and Coding. William G. Chambers. Oxford U Pr, 1985, xvi + 240 pp, $29.95. [ISBN: 0-19-853195-8] This text is built

around the theme of reliable digital communications. Chapter titles include Fourier theory, digital signalling, random variables and noise, signal spaces, error-correcting codes, codes based on finite fields, optical communications, and cipher systems. Includes a good supply of exercises and references. CEC

Applications (Engineering), S*(16-18), P, L*. Mathematical Methods in Engineering. Ed: Glyn A.O. Davies. Handbook of Appl. Math. 5. Wiley, 1984, xiv + 458 pp, $31.95 (P). [ISBN: 0-471-10331-4] One of six guidebooks produced as companion pieces to the six-volume Handbook of Applicable Mathematics. Fourteen chapters by fifteen authors on branches of engineering which particularly rely on the application of mathematical techniques. Brief summaries describe the nature of the subject and "put into context the modelling which is necessary before a mathematical problem emerges." Each chapter has a reading list. References to appropriate core volumes of the Handbook which are designed to elaborate on the mathematical concepts presented in the guidebook. JK

Applications (Physics), T(17-18: 1, 2), S, P. Inequality Problems in Mechanics and Applications: Convex and Nonconvex Energy Functions. P.D. Panagiotopoulos. Birkhauser Boston, 1985, xix + 412 pp, $98.95. [ISBN: 3-7643-3094-5] This book treats boundary value problems of mechanics whose variational forms are inequalities, and emphasizes the relation of inequality problems to convex analysis, the theory of multi-valued differential and integral operators (Clarke's generalized gradient, Warga's derivative container) and optimization theory. Very densely written. MU

Applications (Physics), P. Advances in Nonlinear Waves, Volume II. Ed: Lokenath Debnath. Res. Notes in Math., V. 111. Pitman, 1985, 352 pp, $25.95 (P). [ISBN: 0-273-08648-0] This second volume contains 17 papers which fall into three categories--namely, nonlinear waves in fluids, nonlinear waves in plasmas, and solitons, evolution equations and nonlinear interaction problems. These two volumes arise from a conference held at East Carolina University in 1982. CEC

Applications (Physics), T(15-17), S, L. Introduction to Quantum Mechanics: With Applications to Chemistry. Linus Pauling, E. Bright Wilson, Jr. Dover, 1985, xi + 468 pp, $9.95 (P). [ISBN: 0-486-64871-0] A practical introduction to wave mechanics written in 1935 "for study by men [!] without extensive previous experience with advanced mathematics, such as chemists interested in the subject because of its chemical applications." A thorough, classical, detailed treatment. Unaltered reprint of original McGraw-Hill edition. LAS

Applications (Physics), P. Trends and Developments in the Eighties. Ed: S. Albeverio, Ph. Blanchard. Bielefeld Encounters in Math. & Physics IV & V. World Scientific, 1985, xii + 424 pp, $44. [ISBN: 9971-966-77-8] Written notes of lectures presented at the Centre for Interdisciplinary Research of the University of Bielefeld, July 1982 and September 1983. Contains notes of 17 reviews of current interest. Topics discussed include classical dynamical systems, partial differential equations, stability and instability in quantum systems. AM

Applications (Physics), S(17-18), P. Geometry of Quantum Theory, Second Edition. V.S. Varadarajan. Springer-Verlag, 1985, xviii + 412 pp, $54. [ISBN: 0-387-96124-0] Revision of a two-volume work published by Van Nostrand in 1968 and 1970. Develops quantum mechanics from the viewpoint of von Neumann, in which every physical system is associated with a lattice whose elements represent that system's physically verifiable propositions. Revisions include a number of supplementary notes and references to work published after the First Edition. AM

Applications (Physics), S(14-16). Problems and Solutions in Electromagnetic Theory. C.M. Lerner. Wiley, 1985, x + 614 pp, $34.95 (P). [ISBN: 0-471-88678-5] Collection of 444 problems (with solutions) in electromagnetic theory covering topics including vector analysis, differential geometry, potentials, fields polarization, electric and magnetic boundary value problems, relativity, Green's functions, Maxwell's equations, gauge invariance, conservation laws. Presents problems and solutions clearly with illustrations where appropriate. AM

Applications (Physics), P*. The Mathematics of Combustion. Ed: John D. Buckmaster. Frontiers in Appl. Math. SIAM, 1985, xii + 254 pp, $32.50. [ISBN: 0-89871-053-7] Five chapters by five different authors--introduction to combustion theory, sensitivity analysis of combustion systems, turbulent combustion, detonation in miniature, and finite amplitude waves in combustible gases--each well-referenced. JK

Reviewers

DFA: David F. Appleyard, Carleton; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; RD: Roger Day, St. Olaf; JD-B: John Dyer-Bennet, Carleton; SG: Steven Galovich, Carleton; BH: Bruce Hanson, St. Olaf; PH: Paul Humke, St. Olaf; KK: Kenneth Kaminsky, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; AM: Alan Magnuson, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; LN: Linda Ness, Carleton; AO: Arnold Ostebee, St. Olaf; MS: Michael Schneider, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MT: Michael Tveite, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

For the monkey saddle, however, there is no escaping the lack of smoothness at the umbilic point. A direct computation using the formulas cited above for a surface which is a graph shows that the principal curvatures are given in terms of polar coordinates by

$$\left[-r^5\cos 3\theta \pm 2r\left(1 + r^4 + \frac{r^8}{4}\cos^2 3\theta\right)^{1/2}\right]\Big/(1 + r^4)^{3/2}.$$

As r approaches 0, the principal curvatures λ_1 and λ_2 are asymptotically equal to $2r$ and $-2r$, respectively, and hence they cannot be made smooth at the origin. For small r, the graph of each principal curvature function resembles a cone of revolution, see Fig. 1. This computer-generated figure was produced by T. F. Banchoff of Brown University, and it also appears in [**1**, p. 135].

Of course, the notion of principal curvature generalizes to the case of a higher dimensional hypersurface M^n in E^{n+1}. The shape operator A is again a smooth field of symmetric endomorphisms of the tangent spaces to M. If the principal curvatures are ordered

$$\lambda_1(P) \geq \lambda_2(P) \geq \cdots \geq \lambda_n(P),$$

then each λ_i is a continuous function (see, for example, [**4**, p. 371]). Further, if a continuous principal curvature function λ has constant multiplicity ν on an open subset U of M, then λ and its ν-dimensional distribution of eigenspaces T_λ are both smooth on U (see [**2**]). In addition, the principal distributions are foliations, i.e., they have integral submanifolds (integral curves, in the case $\nu = 1$). For $\nu = 1$, integrability is a consequence of the basic existence theorem for solutions of ordinary differential equations. For $\nu > 1$, one uses the Codazzi equation to show that the integrability conditions of the Frobenius Theorem are satisfied (see [**1**, pp. 139–141]).

Although the multiplicities of the principal curvatures of a hypersurface are always locally constant on a dense open subset [**5**], they need not be globally constant. This possibility of variable multiplicities often presents significant difficulties in obtaining classifications of hypersurfaces.

References

1. T. E. Cecil and P. J. Ryan, Tight and taut immersions of manifolds, Research Notes in Mathematics, No. 107, Pitman, London, 1985.

2. K. Nomizu, Characteristic roots and vectors of a differentiable family of symmetric matrices, Linear and Multilinear Algebra, 2 (1973) 159–162.

3. B. O'Neill, Elementary Differential Geometry, Academic Press, New York, 1966.

4. P. J. Ryan, Homogeneity and some curvature conditions for hypersurfaces, Tôhoku Math. J., 21 (1969) 363–388.

5. D. Singley, Smoothness theorems for the principal curvatures and principal vectors of a hypersurface, Rocky Mountain J. Math., 5 (1975) 135–144.

ZERO DIVISORS IN COMMUTATIVE RINGS

ROBERT GILMER

Department of Mathematics, Florida State University, Tallahassee, FL 32306

For a commutative ring R, we denote by α the cardinality of R and by β the cardinality of the set Y of zero divisors (including zero) in R. We consider the following questions:

1. What conditions are necessary on α and β?
2. If a pair (α, β) of cardinals satisfies these conditions, can we find a corresponding pair (R, Y) so that $|R| = \alpha$ and $|Y| = \beta$?

We will obtain an answer to Question 1 for which the answer to Question 2 is "yes". It turns out that the following cases are easy to resolve: (i) $\beta = 1$, (ii) $\beta = \alpha$, (iii) β is infinite (in which

case $\beta = \alpha$); hence, the most interesting case is that in which $1 < \beta < \alpha < \infty$. In this latter case the characteristic k of R is finite, and in Theorem 4 we give a complete description of the possibilities for k in relation to α and β.

A preliminary answer to Question 1 clearly is that $1 \leqslant \beta \leqslant \alpha$. The extreme case $\beta = 1$ occurs exactly when R is an integral domain. The answer to Question 2 in this case depends on whether α is infinite or finite. Integral domains of infinite cardinality α always exist—for example, the polynomial ring in α indeterminates over the ring of integers. So suppose α is finite. Finite integral domains are fields and a finite field with α elements exists if and only if α is a prime power.

If $\beta > 1$, then there is a useful answer to Question 1:

THEOREM . *Assume that R is a commutative ring of cardinality α having β zero divisors, where $1 < \beta \leqslant \alpha$. Then $\alpha \leqslant \beta^2$ and hence:*

(i) *if β is infinite, then $\alpha = \beta$;*

(ii) *if β is finite, then α is finite.*

Proof. For $x \in R$, the mapping given by $r \to rx$ is a homomorphism of the additive group of R onto the additive group Rx. The kernel is $B = \{r \in R | rx = 0\}$. Hence Rx and R/B are isomorphic as additive groups. It follows that

$$\alpha = |R| = |R/B| \cdot |B| = |Rx| \cdot |B|.$$

If $x \neq 0$ is a zero divisor, then both Rx and B consist of zero divisors of R. Hence $|Rx| \leqslant \beta$ and $|B| \leqslant \beta$, so that $\alpha \leqslant \beta^2$.

For any cardinal $\alpha > 1$, there exists a ring R of cardinality α with $\beta = \alpha$. To see this, let G_α be an abelian group of cardinality α. If the product of any two elements of G_α is defined to be 0, then G_α becomes a commutative ring (called the *zero ring* on G_α) in which every element is obviously a zero divisor. For α infinite, a different type of example with $\alpha = \beta$ is obtained by taking $R = D \times D$, where D is an integral domain of cardinality α; in R, not every element is a zero divisor.

In view of Theorem 1 and the preceding paragraph, we can now restrict consideration to the case where α is finite and $1 < \beta < \alpha$. Note that a finite commutative ring R whose order α is divisible by more than one prime is necessarily *decomposable*—that is, R is a nontrivial direct sum of ideals. In fact, if $|R| = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$ is the prime factorization of $|R|$ and if $R_i = \{x \in R | p_i^{e_i} x = 0\}$, then R_i is an ideal of R of cardinality $p_i^{e_i}$ and $R = R_1 \oplus \cdots \oplus R_k$; the ideal R_i is, of course, just the p_i-primary component of the additive group of R. The relationship between the number of zero divisors in the summands R_i and the number of zero divisors in R is given by

THEOREM 2. *For $1 \leqslant i \leqslant k$, let R_i be a nonzero finite commutative ring with α_i elements and β_i zero divisors. Let $R = R_1 \oplus R_2 \oplus \cdots \oplus R_k$ and let β be the number of zero divisors in R.*

(1) $\beta = \alpha_1 \alpha_2 \cdots \alpha_k - (\alpha_1 - \beta_1)(\alpha_2 - \beta_2) \cdots (\alpha_k - \beta_k)$.

(2) *If $k > 1$, then $\alpha_1 \alpha_2 \cdots \alpha_k = |R| < \beta^2$.*

Proof. (1). We call an element of a commutative ring *regular* if it is not a zero divisor. An element $(x_1, \ldots, x_k)$ of R is regular if and only if x_i is regular in R_i for each i. Hence $(\alpha_1 - \beta_1) \cdots (\alpha_k - \beta_k)$ elements of R are regular, and (1) follows immediately from this fact.

To prove (2), it suffices to consider the case where $k = 2$. Thus, suppose $R = R_1 \oplus R_2$ and $\alpha_1 \geqslant \alpha_2$. The elements $(x, 0)$ and $(0, y)$ are zero divisors in R for each $x \in R_1$, $y \in R_2$. Therefore

$$\beta \geqslant \alpha_1 + \alpha_2 - 1 \geqslant \alpha_1 + 1 \quad \text{and} \quad \beta^2 \geqslant (\alpha_1 + 1)^2 > \alpha_1 \alpha_2.$$

N. Ganesan [1] has investigated relations between α and β when β is finite and $1 < \beta < \alpha$; in particular, he proved (ii) of Theorem 1. One of the main questions considered in [1] is whether there exists, for each $n > 1$, a commutative ring with n^2 elements and exactly n zero divisors. For

n prime, Ganesan observed that Z/n^2Z is such a ring. The following corollary provides a complete answer.

COROLLARY. *A finite commutative ring of order n^2 having $n > 1$ zero divisors exists if and only if n is a prime power. Such a ring is indecomposable.*

Proof. Suppose $n = p^w$ with p prime. Let F be a field with p^w elements and let $R = F[X]/(X^2)$. Then $|R| = p^{2w}$ and $(X)/(X^2)$, the set of zero divisors of R, has cardinality p^w. More generally, if k is any positive integer, then $F[X]/(X^{k+1})$ is a ring of cardinality $p^{w(k+1)}$ with p^{wk} zero divisors. Conversely, suppose that R has order n^2 and n zero divisors. Then R is indecomposable, since the decomposability of R would contradict part (2) of Theorem 2. Furthermore n^2 (and hence n) must be a prime power because otherwise R would be decomposable.

Returning to the original questions, suppose R is a finite commutative ring with α elements and β zero divisors, where $1 < \beta < \alpha$. To determine other relations that must exist between α and β, we exploit more of the known theory concerning the decomposability of R. To wit, since R is finite and contains a regular element, R has an identity and every regular element of R is invertible (to prove this, use the fact that $Rx = R$ for any regular element $x \in R$). Moreover, the finiteness of R implies that R is expressible as a finite direct sum $R_1 \oplus \cdots \oplus R_n$, where each R_i is nonzero and indecomposable and has an identity. Theorem 2 then shows that β is determined by the pairs (α_i, β_i), where R_i has α_i elements and β_i zero divisors.

Indecomposability of R_i implies that α_i is a prime power. More importantly, it implies that R_i is *local*, meaning that the set M of nonunits of R_i is the unique maximal ideal of R_i. To prove that M is an ideal, we first note that each $x \in M$ is nilpotent (that is, $x^t = 0$ for some $t \geqslant 1$). If not, then the finiteness of R_i implies that the powers of x are not all distinct; it follows that some power of x, say $e = x^s$, has the property that $e^2 = e \neq 0$. This, in turn, implies that R_i is decomposable [**2**, Exercise 23, p. 135], a contradiction. So every nonunit is nilpotent; the nilpotent elements in any commutative ring form an ideal [**2**, Exercise 1, p. 133]. The maximality of M is immediate since every proper ideal of R_i consists of nonunits.

Because the zero divisors of R_i form an ideal, it follows that $\beta_i = |M|$ divides $\alpha_i = |R|$ in this case; therefore β_i is also a prime power. Hence the possibilities for β can be determined from the next result.

THEOREM 3. *Assume that p is prime and s and t are integers such that $0 \leqslant s < t$. There exists an indecomposable ring of order p^t having p^s zero divisors if and only if $t - s$ divides s.*

Proof. Assume first that $t - s = w$ divides s—say, $s = kw$ and $t = (k+1)w$. Then as previously observed, $F[X]/(X^{k+1})$, where F is a field with p^w elements, is an indecomposable ring of order p^t having p^s zero divisors.

Conversely, if R is indecomposable of order p^t with p^s zero divisors, then R has a unique maximal ideal M and $|M| = p^s$. Consider R as a module over itself. Finiteness of R implies that it has a composition series

$$R = I_0 > I_1 = M > I_2 > \cdots > I_k > I_{k+1} = (0)$$

[**2**, pp. 375–76]; the submodules I_j are ideals of R. Since there are no ideals between I_j and I_{j+1}, then I_j/I_{j+1} is a simple module, and therefore isomorphic to R/M for each j [**2**, Theorem 1.3, p. 417]. Therefore

$$|I_j/I_{j+1}| = |R/M| = p^{t-s};$$

by induction we obtain $|I_{k-j+1}| = p^{j(t-s)}$ for each j. Hence

$$p^s = |I_1| = p^{k(t-s)}$$

and $t - s$ divides s as asserted.

We remark that an indecomposable commutative ring of order $p^{(k+1)w}$ with p^{kw} zero divisors

is determined up to isomorphism only if $k = 0$. To see this, we note that if $k > 1$, then

$$R_1 = F[X]/(X^{k+1}) \quad \text{and} \quad R_2 = F[X_1,\dots,X_k]/(X_1,\dots,X_k)^2,$$

where F is a field of order p^w, are local rings of order $p^{(k+1)w}$ with p^{kw} zero divisors; $R_1 \not\cong R_2$ since $x^2 = 0$ for each zero divisor x of R_2, while this condition fails in R_1. If $k = 1$, we again take $R_1 = F[X]/(X^2)$ and we take $R_3 = Z[X]/(p^2, f)$, where $f \in Z[X]$ is monic of degree w and is irreducible modulo p. It is straightforward to check that R_3 is local with maximal ideal $(p, f)/(p^2, f)$, has order p^{2w}, and has p^w zero divisors. On the other hand, char $R_1 = p$ and char $R_3 = p^2$, so $R_1 \not\cong R_3$.

To illustrate the use of Theorems 2 and 3, we present three tables listing values of β which occur for selected values of α. Since $\beta = \alpha$ is always a possibility, the tables consider the case where $1 \leqslant \beta < \alpha$, which coincides with the case where R has an identity element. Table 1 lists possibilities for β when R is indecomposable of order p^t, $1 \leqslant t \leqslant 6$; it is obtained using Theorem 3. Table 2 lists values of β when the prime factorization of α has one of four elementary forms. Finally, values of β that occur for $1 < \alpha < 32$, and which are not covered by Table 2, are listed in Table 3.

TABLE 1. Possibilities for β if R is indecomposable.

α	β
p	1
p^2	$1, p$
p^3	$1, p^2$
p^4	$1, p^2, p^3$
p^5	$1, p^4$
p^6	$1, p^3, p^4, p^5$

TABLE 2. α has a simple prime factorization (p and q denote distinct primes).

α	β
p	1
p^2	$1, p, 2p - 1$
p^3	$1, p^2, p^2 + p - 1, 2p^2 - p, 3p^2 - 3p + 1$
pq	$p + q - 1$

TABLE 3. Six small values of α.

α	β
12	6, 8, 10
16	1, 4, 7, 8, 9, 10, 12, 13, 14, 15
18	10, 12, 14
20	8, 12, 16
24	10, 16, 18, 20, 22
28	10, 16, 22
30	22

We justify only the line for $\alpha = p^3$ in Table 2. Thus, suppose R is a commutative ring with identity of order p^3. If R is indecomposable, then Theorem 3 shows that $\beta = 1$ or $\beta = p^2$. Otherwise, R is the direct sum of nonzero indecomposable rings, and there are two possibilities for the orders of the summands, as follows:

Case 1. $R = R_1 \oplus R_2$, where $|R_1| = p$ and $|R_2| = p^2$.
Case 2. $R = R_1 \oplus R_1 \oplus R_1$.

The ring R_1 is a field, and Theorem 3 shows that R_2 has 1 or p zero divisors. Theorem 2 then shows that R has

$$p^3 - (p - 1)^3 = 3p^2 - 3p + 1$$

zero divisors in Case 2, and in Case 1, R has either

$$p^3 - (p-1)(p^2-1) = p^2 + p - 1$$

or

$$p^3 - (p-1)(p^2-p) = 2p^2 - p$$

zero divisors.

Suppose R is a commutative ring of order α having $\beta < \alpha$ zero divisors. Ganesan considers possibilities for the characteristic k of R. He shows [**1**, p. 217] that the conditions $k|\alpha$ and $\phi(k) \leqslant (\alpha - \beta) \leqslant \beta^2 - \beta$ are necessary, and he asks if they are also sufficient. This paper shows that the answer is negative, as the interested reader may verify. What then are the possibilities for k? By decomposing R into indecomposable summands, we can reduce to the case where R is as in Theorem 3—an indecomposable ring. Theorem 4 then resolves the question of possible characteristics of R. We remark that for finite rings R, the conditions of being indecomposable and being local are equivalent. We have already shown that an indecomposable finite ring is local. The converse holds because in any nontrivial direct sum $R = R_1 \oplus R_2$ of rings with identity, the identity element of R is the sum of two nonunits.

THEOREM 4. *Suppose R is a local ring of order p^{mw} having $p^{(m-1)w}$ zero divisors. The characteristic of R is p^v for some $v \leqslant m$. Conversely, if m, v, and w are positive integers with $v \leqslant m$, then there exists a local ring S of order p^{mw} such that* char $S = p^v$ *and S has $p^{(m-1)w}$ zero divisors.*

Proof. If the characteristic of R is p^v, then v is the smallest positive integer g such that $p^g = 0$. If $v = 1$, we're finished. Otherwise, the sequence of ideals $pR \supseteq p^2R \supseteq \cdots \supseteq p^{v-1}R \supseteq (0)$ properly descends, for if $p^iR = p^{i+1}R$ for some $i < v$, then

$$p^iR = p^{i+1}R = \cdots = p^vR = (0),$$

a contradiction. The series $pR > p^2R > \cdots > p^{v-1}R > (0)$ can be refined to a composition series for pR of length $u \geqslant v - 1$. The proof of Theorem 3 then shows that $|pR| = p^{uw} \leqslant p^{(m-1)w}$. Hence $v - 1 \leqslant m - 1$ and $v \leqslant m$.

For the converse, divide m by v using the division algorithm: $m = hv + i$, where $0 \leqslant i \leqslant v - 1$. Let

$$T = Z[X]/(p^v, f^{h+1}),$$

where f is a monic polynomial of degree w that is irreducible modulo p. Then T is a local ring of order $p^{vw(h+1)}$, with maximal ideal

$$M = (p, f)/(p^v, f^{h+1})$$

of order $p^{vwh+(v-1)w}$. Moreover, char$T = p^v$. Let

$$I_j = (p^v, p^jf^h, f^{h+1})/(p^v, f^{h+1})$$

for $0 \leqslant j \leqslant v - 1$. Then $MI_j \subseteq I_{j+1}$ for each j, $MI_{v-1} = (0)$, and these facts can be used in showing that

$$(p^v, f^h)/(p^v, f^{h+1}) = I_0 > I_1 > \cdots > I_{v-1} > (0)$$

is a composition series for I_0. From the proof of Theorem 3, it follows that $|I_j| = p^{w(v-j)}$ for each j. Let

$$S = T/I_i \simeq Z[X]/(p^v, p^if^h, f^{h+1}).$$

Then S is a local ring with maximal ideal

$$M/I_i \simeq (p, f)/(p^v, p^if^h, f^{h+1}).$$

The cardinality of S is

$$|T|/|I_i| = p^{vw(h+1)}/p^{w(v-i)} = p^{w(vh+i)} = p^{wm},$$

while M/I_i has cardinality $p^{w(m-1)}$. Since S is a homomorphic image of T, its characteristic divides p^v, the characteristic of T. Because

$$S/(I_0/I_i) \simeq T/I_0 \simeq Z[X]/(p^v, f^h)$$

is a homomorphic image of S, similar reasoning shows that char S is a multiple of char$(T/I_0) = p^v$. Therefore S has characteristic p^v, and this completes the proof of Theorem 4.

References

1. N. Ganesan, Properties of rings with a finite number of zero divisors, Math. Ann., 157(1964) 215–218.
2. T. W. Hungerford, Algebra, Holt, Rinehart and Winston, New York, 1974; reprint edition, Springer-Verlag, New York, 1981.

HOW TO COMPUTE THE SERIES EXPANSIONS OF SEC X AND TAN X

M. D. ATKINSON

School of Computer Science, Carleton University, Ottawa, Canada K1S 5B6

Of the six standard trigonometric functions cos, sin, tan, sec, cosec and cot only the first two have easy to remember series expansions. Expansions for the others exist but the coefficients involve Bernoulli and Euler numbers and do not display any evident regularity. In this note we shall give a remarkably simple method for obtaining the coefficients of the series expansions of $\sec x$ and $\tan x$. It requires only addition operations and is well suited to both hand calculations and computer calculations.

Suppose that

$$\sec x = \sum_{n=0}^{\infty} s_n x^n/n!$$

and

$$\tan x = \sum_{n=0}^{\infty} t_n x^n/n!.$$

Then, since $\sec x$ is an even function and $\tan x$ is an odd function, s_n is non-zero only for even values of n and t_n is non-zero only for odd values. These non-zero values may be read off from the following triangular table.

						1						
					0		1					
				1		1		0				
			0		1		2		2			
		5		5		4		2		0		
	0		5		10		14		16		16	
61		61		56		46		32		16		0

. . . .

The secant coefficients $1,0,1,0,5,0,61,\ldots$ appear on the left arm of the triangle and the tangent coefficients $1,0,2,0,16,0,\ldots$ appear on the right arm. The reader might enjoy deducing the simple rule whereby the triangle is constructed before continuing to the next paragraph.

The triangle has a single 1 in its first row (row 0). Each subsequent row is formed from the partial sums of the previous row; for odd numbered rows the partial summation is from left to right and for even numbered rows it is from right to left (in both cases beginning with a zero).

To justify this rule we begin by recalling the formula for multiplying two exponential

generating functions, namely

$$\sum a_n x^n/n! \sum b_n x^n/n! = \sum c_n x^n/n! \quad \text{where} \quad c_n = \sum a_r b_{n-r}\binom{n}{r}.$$

Applying this to the identities

$$\cos x \sec x = 1 \quad \text{and} \quad \sin x \sec x = \tan x$$

and using the standard expansions of $\cos x$ and $\sin x$ we obtain the recurrences

$$s_n - \binom{n}{2}s_{n-2} + \binom{n}{4}s_{n-4} - \cdots = 0 \quad \text{for even } n > 0,$$

$$t_n - \binom{n}{1}s_{n-1} + \binom{n}{3}s_{n-3} - \cdots = 0 \quad \text{for odd } n > 1,$$

together with $s_0 = 1$ and $t_1 = 1$. (Of course these equations also may be used to compute s_n and t_n but considerably more effort is required.)

Now let $z_{nr} (n \geqslant 0,\ r \geqslant 0)$ be the rth element in the nth row of the triangle. By definition

$$z_{n0} = 0 \quad \text{and} \quad z_{n,r+1} = z_{nr} + z_{n-1,r} \quad \text{if } n \text{ is odd},$$

and

$$z_{nn} = 0 \quad \text{and} \quad z_{nr} = z_{n,r+1} + z_{n-1,r} \quad \text{if } n \text{ is even, } n > 0.$$

Consequently, using the forward difference operator Δ (applied to the second subscript) we have

$$z_{n-1,r} = \Delta z_{nr} \quad \text{if } n \text{ is odd},$$

and

$$z_{n-1,r} = -\Delta z_{nr} \quad \text{if } n \text{ is even}.$$

Therefore

$$\Delta^k z_{nr} = (-1)^{n+[k/2]-1} z_{n-k,r}. \tag{1}$$

By the summation formula [2, p. 121] of the difference calculus

$$\sum_{r=0}^{n-1} z_{n-1,r} = \binom{n}{1}z_{n-1,0} + \binom{n}{2}\Delta z_{n-1,0} + \binom{n}{3}\Delta^2 z_{n-1,0} + \cdots$$

$$= \begin{cases} \binom{n}{2}z_{n-2,0} - \binom{n}{4}z_{n-4,0} + \cdots & \text{if } n \text{ is even, } n > 0 \\ \binom{n}{1}z_{n-1,0} - \binom{n}{3}z_{n-3,0} + \cdots & \text{if } n \text{ is odd,} \end{cases}$$

(using equation (1) and the fact that $z_{m0} = 0$ if m is odd).

However, the left-hand side of this equation is the sum of all the elements in the $(n-1)$th row of the triangle and so it is equal to z_{n0} if n is even, and equal to z_{nn} if n is odd. This means that the two sequences $(z_{00}, z_{20}, z_{40}, \ldots)$ and $(z_{11}, z_{33}, z_{55}, \ldots)$ satisfy the same recurrence (respectively) as $(s_0, s_2, s_4, \ldots)$ and $(t_1, t_3, t_5 \ldots)$. Since $z_{00} = s_0$ and $z_{11} = t_1$, it follows that $z_{n0} = s_n$ for all even n and $z_{nn} = t_n$ for all odd n, thereby justifying the method.

The series expansion of $\cot x$ and $\operatorname{cosec} x$ can also be obtained since if

$$\cot x = x^{-1} + \sum c_n x^n/n!,$$

and

$$\operatorname{cosec} x = x^{-1} + \sum e_n x^n/n!,$$

we have [1]

$$c_n = -t_n/(2^{n+1} - 1) \quad \text{and} \quad e_n = t_n(1 - 2^{-n})\big/(2^{n+1} - 1).$$

References

1. M. Abramowitz and I. A. Stegun, Handbook of Mathematical Functions, Dover Publications, New York, 1968.
2. M. R. Spiegel, Calculus of Finite Differences and Difference Equations, Schaum Outline Series, McGraw-Hill, New York, 1971.

---

168. MISCELLANEA

Computers are icumen in

There was a time when teachers
Could just get up and talk.
Slateboards were invented;
You wrote on them with chalk.
At least this slowed the lecture
So the audience could try
To take some notes on everything
Before it passed them by.
But slate was too expensive,
So substitutes arose
That scattered dust on everything
But mostly on our clothes.
(A kind that's best forgotten—
It died while it was new—
The yellow board with purple chalk
That turned your fingers blue.)
The overhead projector
Seems to help, as well it may,
As long as you can make enough
Transparencies each day.
But I have seen the future,
And I don't like what I see:
Computers in the classroom,
They come too fast for me.
There's lots and lots of labor,
As far as I can hear,
Creating complex programs
To make simple topics clear.
There might be an alternative
We shouldn't overlook:
What's wrong with telling students
That they have to read the book?

Ralph P. Boas
Department of Mathematics
Northwestern University
Evanston, IL 60201

AN ELEMENTARY APPROACH TO THE FUNCTIONAL CALCULUS FOR MATRICES

YASUHIKO IKEBE AND TOSHIYUKI INAGAKI
Institute of Information Sciences and Electronics, University of Tsukuba, Ibaraki 305, Japan

Let A be a square matrix of complex numbers. Let $(\lambda_1 - \lambda)^{m_1} \cdots (\lambda_r - \lambda)^{m_r}$ be the minimal polynomial of A, where $\lambda_1, \dots, \lambda_r$ is a complete enumeration of the distinct eigenvalues of A and m_j $(j = 1, \dots, r)$ equals the largest order of the Jordan blocks with eigenvalue λ_j. Let f be a function analytic on a neighborhood of $\{\lambda_1, \dots, \lambda_r\}$. It is known that a unique polynomial g of degree at most $m_1 + \cdots + m_r - 1$ exists that satisfies the interpolation conditions

$$g(\lambda_j) = f(\lambda_j), \dots, g^{(m_j-1)}(\lambda_j) = f^{(m_j-1)}(\lambda_j), \quad j = 1, \dots, r.$$

The polynomial g, which depends on f, is called the Lagrange-Sylvester interpolation polynomial [**1**, Chapter 5] or the Hermite interpolation polynomial [**3**, Chapter 1] for f on the spectrum of A. A matrix $f(A)$ may then be defined by setting $f(A) = g(A)$, where the right side means the polynomial in A that is obtained by substituting A for z in $g(z)$; see [**1**, Chapter 5].

The aim of this note is to give an alternative and equivalent definition of $f(A)$ that is more intuitive and transparent, using the Jordan canonical form of A in a more direct way. A key to this is given by a homomorphism from an algebra of functions to a commutative algebra of upper-triangular matrices, to be stated shortly. Our purpose in presenting this material, which is known to many researchers in the subject, is to make this method more widely known among teachers of linear algebra. With our approach, we can dispense with the use of interpolation polynomials, thus making it possible to present the functional calculus for matrices to students with a modest background in matrix theory in a more direct and understandable form. Theorem 4 guarantees that the two definitions of $f(A)$ (namely, Gantmacher's and ours) agree. Our approach leads in a natural way to Dunford's integral representation of $f(A)$ (see Dunford & Schwartz [**2**, Chapter 7]), thus filling a gap between the elementary and advanced theories of functions of a single operator; see Theorem 5.

We will give only the definitions and the statements of the theorems that lead to a definition of $f(A)$. Except for Theorem 5, the proofs are left as exercises for the reader.

Let f be analytic on a neighborhood of $\{\lambda_1, \dots, \lambda_r\}$. Let n denote a positive integer. By the symbol $f^*(z)$, or simply f^*, we mean the n by n upper-triangular matrix defined by:

$$f^*(z) = \begin{pmatrix} f(z) & f'(z) & f''(z)/2! & \dots & f^{(n-1)}(z)/(n-1)! \\ & f(z) & f'(z) & \ddots & \vdots \\ & & \ddots & \ddots & f'(z) \\ 0 & & & & f(z) \end{pmatrix}$$

This matrix appears also in the Gantmacher's treatment [**1**, Chapter 5], though in different contexts.

We note that for $f(z) = z$, we have

$$f^*(\lambda) = \begin{pmatrix} \lambda & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda \end{pmatrix} = J,$$

i.e., a single Jordan block with eigenvalue λ. If $f(z) \equiv c$ (constant), then $f^*(\lambda) = cI$, where I denotes the identity matrix.

The following key theorems, easily verifiable, state that the map $f \to f^*$ is a homomorphism

from the algebra of functions analytic on a neighborhood of $\{\lambda_1, \dots, \lambda_r\}$ to a commutative algebra of upper-triangular matrices.

THEOREM 1.
(1) $(f+g)^* = f^* + g^*$,
(2) $(cf)^* = cf^*$, *where c is a constant*,
(3) $(fg)^* = f^*g^* = g^*f^*$,
(4) $(f/g)^* = f^*(g^*)^{-1} = (g^*)^{-1}f^*$ *if* $g(z) \neq 0$,
(5) $(1/g)^* = (g^*)^{-1}$ *if* $g(z) \neq 0$.

As an immediate corollary to Theorem 1 we have the following theorem.

THEOREM 2. *Let J be a single Jordan block with eigenvalue* λ. *Let f be a rational function not having a pole at* λ, *and let p and q be co-prime polynomials such that* $f = p/q$. *Then*

$$f^*(\lambda) = p(J)[q(J)]^{-1} = [q(J)]^{-1}p(J).$$

The next theorem states that a similar result holds for an infinite power series.

THEOREM 3. *Let J be a single Jordan block with eigenvalue* λ. *Let* $f(z) = a_0 + a_1 z + \cdots$ *be an infinite power series whose radius of convergence is strictly greater than* $|\lambda|$. *Then*

$$f^*(\lambda) = a_0 I + a_1 J + \cdots .$$

The Jordan canonical form of a square matrix A is given by

$$V^{-1}AV = \operatorname{diag}[J_1, \dots, J_m],$$

where V is an invertible matrix and the right side denotes the block diagonal matrix with the diagonal blocks $J_1, \dots, J_m$. Here J_i denotes a single Jordan block with eigenvalue μ_i ($i = 1, \dots, m$). The μ_i are not necessarily distinct. If f is analytic on a neighborhood of the eigenvalues $\{\mu_1, \dots, \mu_m\}$ of A, then we can define $f(A)$ by

$$V^{-1}f(A)V = \operatorname{diag}[f^*(\mu_1), \dots, f^*(\mu_m)].$$

Theorems 2 and 3 show that $f(A)$ is what we expect if f is either rational or a power series.

As further applications of foregoing theorems, we obtain the following two theorems. The first one gives a necessary and sufficient condition for equality of $f(A)$ and $g(A)$. The second one leads us to the Dunford's integral representation of $f(A)$.

THEOREM 4 (Identity theorem). *Let* $\lambda_1, \dots, \lambda_r$ *be distinct eigenvalues of a square matrix A. Let f and g be analytic on a neighborhood of* $\{\lambda_1, \dots, \lambda_r\}$. *Then* $f(A) = g(A)$ *if and only if*

$$f^{(i)}(\lambda_j) = g^{(i)}(\lambda_j), \quad i = 0, 1, \dots, m_j - 1, \quad j = 1, \dots, r,$$

where m_j *denotes the largest order of the Jordan blocks with eigenvalue* λ_j.

THEOREM 5 (Dunford's integral representation of $f(A)$). *Let C be a simple closed curve that encloses in its interior every eigenvalue of a square matrix A. Let f be analytic on C and in the interior of C. Then*

$$f(A) = \frac{1}{2\pi i}\int_C f(t)(tI - A)^{-1}\,dt,$$

where the right side means, by definition, the elementwise integration.

Proof. It is sufficient to prove the above equation for a single Jordan block J of order k with eigenvalue λ_j. We have

$$f(J) = f^*(\lambda_j) \quad \text{(by Theorems 2 and 3)}$$

$$= \frac{1}{2\pi i}\int_C f(t)\begin{pmatrix} \frac{1}{t-\lambda_j} & \frac{1}{(t-\lambda_j)^2} & \cdots & \frac{1}{(t-\lambda_j)^k} \\ & \ddots & \ddots & \vdots \\ & & \ddots & \frac{1}{(t-\lambda_j)^2} \\ 0 & & & \frac{1}{t-\lambda_j} \end{pmatrix} dt$$

$$= \frac{1}{2\pi i}\int_C f(t)(tI - J)^{-1}\, dt.$$

References

1. F. R. Gantmacher, The Theory of Matrices, Vol. 1, Chelsea Publishing Company, New York, 1959.
2. N. Dunford and J. T. Schwartz, Linear Operators, Part I: General Theory, Wiley Interscience, New York, 1958.
3. B. Wendroff, Theoretical Numerical Analysis, Academic Press, New York, 1966.

THE TEACHING OF MATHEMATICS

Edited by Mary R. Wardrop and Robert F. Wardrop
For instructions about submitting material for publication in this department see the inside front cover.

EIGENVALUES AND EIGENVECTORS OF "N-MATRICES"

Samuel Councilman
Department of Mathematics and Computer Science, California State University, Long Beach, CA 90840

A standard topic in Elementary Linear Algebra is the determination of the *eigenvalues* and *eigenvectors* of square matrices: Given an $n \times n$ (real) matrix A, find all (real) numbers λ and all vectors $X = [x_1, x_2, \ldots, x_n]$ such that

$$(*) \qquad AX^T = \lambda X^T,$$

where X^T denotes the transpose of X.

The (real) numbers λ for which equation $(*)$ holds are called the *characteristic values* or *eigenvalues* of A and the corresponding vectors X are called the *characteristic vectors* or *eigenvectors* of A. The eigenvalues of matrix A are precisely the roots (zeros) of the *characteristic polynomial of* A, $p_A(\lambda) =$ determinant $(A - \lambda I_n)$. For each eigenvalue λ_i the corresponding eigenvectors are found by solving the linear system $(A - \lambda_i I_n)X^T = 0$.

The main result in this subject area is the well-known

THEOREM. *An $n \times n$ matrix A is "diagonalizable" if and only if A has n linearly independent eigenvectors.*

The essence of this theorem is that whenever A has n linearly independent eigenvectors, they are used as the column vectors of a non-singular matrix B and then $B^{-1}AB$ is the diagonal matrix having the eigenvalues of A as the diagonal entries. Thus A is "diagonalized." If A does not have

n linearly independent eigenvectors one cannot form the matrix B and matrix A is *not* "diagonalizable."

When teaching the topics of eigenvalues, eigenvectors and diagonalizability, it is quite desirable to have an abundant supply of examples and problems which illustrate all possible situations that can occur with respect to distinct or multiple eigenvalues and the existence of n or fewer linearly independent eigenvectors.

Most linear algebra texts do a rather poor job in this regard. Either they fail to cover all possible cases or, if they do, their examples and problems suffer from one or the other of two possible defects:

(i) Computing the characteristic polynomial $\det(A - \lambda I)$ is very tedious and finding its roots is somewhat difficult, or

(ii) matters are oversimplified by using only upper or lower triangular, or even diagonal, matrices.

The drawback in case (ii) is that the eigenvalues are the diagonal entries, so the student is not required to compute the characteristic polynomial and find its roots. Furthermore, the system $(A - \lambda_i I)X^T = 0$ is already in "echelon form," so students get little practice in using Gaussian elimination.

Fortunately matrices of the form

$$A = \begin{bmatrix} a_{11} & 0 & 0 & 0 & \cdots & a_{1n} \\ a_{21} & a_{22} & 0 & 0 & \cdots & a_{2n} \\ a_{31} & 0 & a_{33} & 0 & \cdots & a_{3n} \\ \vdots & \vdots & & & & \vdots \\ a_{n1} & 0 & 0 & 0 & \cdots & a_{nn} \end{bmatrix}$$

offer an exceptional source of computationally convenient eigenvalue-eigenvector problems which have neither of the aforementioned defects. In view of the pattern of (possible) non-zero entries, one might call such matrices "N-matrices."

Because of the zeros in columns 2 through $(n-1)$ of an N-matrix, computation of its characteristic polynomial $\det(A - \lambda I_n)$ is very easy. Expanding by cofactors of any one of columns 2 through $(n-1)$, one finds that the minor of the non-zero element is itself the determinant of an N-submatrix. This minor in turn can be expanded by cofactors of one of its "special columns." In fact, by repeated such expansion one obtains the simple

$$\begin{aligned} \det(A - \lambda I_n) &= [a_{22} - \lambda][a_{33} - \lambda] \cdots [a_{(n-1)(n-1)} - \lambda] \begin{vmatrix} a_{11} - \lambda & a_{1n} \\ a_{n1} & a_{nn} - \lambda \end{vmatrix} \\ &= [a_{22} - \lambda][a_{33} - \lambda] \cdots [a_{(n-1)(n-1)} - \lambda] \\ &\quad \times [\lambda^2 - (a_{11} + a_{nn})\lambda + (a_{11}a_{nn} - a_{1n}a_{n1})]. \end{aligned}$$

By proper selection of the elements in an N-matrix one can obtain a great variety of examples and problems. The following details demonstrate exactly how N-matrices can be used to illustrate all possible eigenvalue-eigenvector situations for 3×3 matrices.

N-matrices of the form $\begin{bmatrix} a & 0 & b \\ c & d & c \\ b & 0 & a \end{bmatrix}$ have characteristic polynomial

$$\det(A - \lambda I_3) = [d - \lambda][\lambda - (a + b)][\lambda - (a - b)]$$

and thus eigenvalues d, $a + b$ and $a - b$. The numbers a, b and d may be chosen so as to yield the following possible cases:

(1) If a, b and d are selected so that there is *no double eigenvalue*, one obtains three linearly independent eigenvectors and the matrix is diagonalizable. Namely,

eigenvalue d has corresponding eigenvectors $k[0,1,0]$,

eigenvalue $a+b$ has corresponding eigenvectors $k\left[1, \dfrac{2c}{(a+b)-d}, 1\right]$,

eigenvalue $a-b$ has corresponding eigenvectors $k[-1,0,1]$,

where k denotes any real number. One possible matrix B is then

$$\begin{bmatrix} 0 & 1 & -1 \\ 1 & x & 0 \\ 0 & 1 & 1 \end{bmatrix} \text{ with } x = \frac{2c}{(a+b)-d}, \quad \text{and} \quad B^{-1}AB \quad \text{is just} \quad \begin{bmatrix} d & 0 & 0 \\ 0 & a+b & 0 \\ 0 & 0 & a-b \end{bmatrix}.$$

(2) If $c \neq 0$ and a, b and d are selected so that there *is a double eigenvalue* $d = a+b$, one obtains only two linearly independent eigenvectors and the matrix is not diagonalizable. This occurs because the double eigenvalue $d = a+b$ has corresponding eigenvectors $k[0,1,0]$.

However,

(3) if a, b and d are selected so that $d = a-b$ *is a double eigenvalue*, then one again obtains three linearly independent eigenvectors and the matrix is diagonalizable. This happens because the double eigenvalue $d = a-b$ has corresponding eigenvectors $k_1[-1,0,1] + k_2[0,1,0]$.

Thus one possible matrix B is $\begin{bmatrix} 1 & -1 & 0 \\ x & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$ with $x = \dfrac{2c}{(a+b)-d}$, and $B^{-1}AB = \begin{bmatrix} a+b & 0 & 0 \\ 0 & d & 0 \\ 0 & 0 & d \end{bmatrix}$.

It is interesting to note that if $c = 0$ in case (2), one again obtains a diagonalizable matrix! This time the double eigenvalue $d = a+b$ has corresponding eigenvectors $k_1[1,0,1] + k_2[0,1,0]$.

N matrices of the form $\begin{bmatrix} d+q & 0 & q \\ c & d & e \\ -q & 0 & d-q \end{bmatrix}$ have characteristic polynomial $\det(A - \lambda I_3) = -(\lambda - d)^3$ and so have d as a *triple eigenvalue*. In this case, by judicious choice of the entries c and e, one can obtain either one or two linearly independent eigenvectors. To be specific,

(4) if $c \neq e$ the eigenvalue d has corresponding eigenvectors $k[0,1,0]$, whereas

(5) if $c = e$ the corresponding eigenvectors are $k_1[0,1,0] + k_2[-1,0,1]$.

(6) For the sake of completeness one might observe that the case $q = 0$, $c = e = 0$ yields the diagonal matrix $\begin{bmatrix} d & 0 & 0 \\ 0 & d & 0 \\ 0 & 0 & d \end{bmatrix}$ which has the triple eigenvalue $\lambda = d$ and three linearly independent eigenvectors $k_1[1,0,0] + k_2[0,1,0] + k_3[0,0,1]$.

It is obvious how the basic results (1)–(6) can be incorporated into 4×4, 5×5, etc., matrices so as to obtain any desired situation. For example, the matrix

$$\begin{bmatrix} a & 0 & 0 & b \\ c_2 & d & 0 & c_2 \\ c_3 & 0 & e & c_3 \\ b & 0 & 0 & a \end{bmatrix}$$

has *four distinct eigenvalues* if $d \neq e$ and neither equals $a+b$ or $a-b$, but it has *two double eigenvalues* if, say, $d = a+b$ and $e = a-b$.

Those desiring an even greater diversity of examples and problems may turn to *transposes* of N-matrices, that is matrices of the form

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & 0 & \cdots & 0 \\ 0 & 0 & a_{33} & \cdots & 0 \\ \vdots & \vdots & & & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix}.$$

These matrices have the same easily computable characteristic polynomials as N-matrices and their entries may be selected so as to yield results analogous to the N-matrix cases. For 3×3 matrices one obtains the same six cases (1)–(6) discussed above, but in each of the cases there are slight differences in the *form of the eigenvectors*.

A TRULY ELEMENTARY APPROACH TO THE BOUNDED CONVERGENCE THEOREM

JONATHAN W. LEWIN
Department of Mathematics, Kennesaw College, Marietta, GA 30061

The bounded convergence theorem follows trivially from the Lebesgue dominated convergence theorem, but at the level of an introductory course in analysis, when the Riemann integral is being studied, how hard is the bounded convergence theorem? For an answer, we might look at Bartle and Sherbert [**2**], page 203: *The proof of this result is quite delicate and will be omitted*. Or we might look at Apostol [**1**], page 228: *The proof of Arzela's theorem is considerably more difficult than ... and will not be given here*. Walter Rudin in [**4**] ignores the theorem altogether in his chapter on Riemann integration, presenting it only as a corollary to the Lebesgue dominated convergence theorem several chapters later, and in [**5**], in an interesting problem in Chapter Two, Rudin refers his readers to [**3**]. In [**3**], Eberlein does present a proof which from some points of view is elementary. Certainly, his proof does not require any notions of measurability, but it is hardly elementary from the point of view of a student who is first learning the Riemann integral. So the answer to the above question seems to be: very hard! But this is not so. In this paper, we present the proof of the bounded convergence theorem in a truly elementary setting, and in such a way that it could be included for the first time in an introductory course.

We begin by defining an elementary set. A bounded subset E of R is said to be *elementary* if E is a finite union of bounded intervals, or equivalently, if χ_E is a step function. One can define the Lebesgue measure $m(E)$ of an elementary set E to be $\int_a^b \chi_E$, where $[a, b]$ is an interval including E, and one can show simply that on the family of elementary sets (which is closed under union, intersection and differences), Lebesgue measure is finitely additive and finitely subadditive. Given a Riemann integrable function f on an interval $[a, b]$, and an elementary subset E of $[a, b]$, we define $\int_E f = \int_a^b f\chi_E$. If E and F are mutually disjoint elementary sets, then one may show easily that $\int_{E \cup F} f = \int_E f + \int_F f$, and if $|f(x)| \leqslant K$ for every point x in E, then $|\int_E f| \leqslant Km(E)$. One may also prove simply that if E is an elementary set and $\varepsilon > 0$, then one can find a closed elementary subset H of E such that $m(H) > m(E) - \varepsilon$.

LEMMA. *Suppose (A_n) is a contracting sequence of bounded subsets of R, with an empty intersection. For each n, define*

$$\alpha_n = \sup\{ m(E) \,|\, E \text{ is an elementary subset of } A_n \}.$$

Then $\alpha_n \to 0$ as $n \to \infty$.

Proof. The sequence (α_n) is clearly decreasing. Now, to obtain a contradiction, assume that this sequence does not converge to 0, and choose $\delta > 0$, such that $\alpha_n > \delta$ for all n. For each n, choose a closed elementary subset E_n of A_n such that

$$m(E_n) > \alpha_n - \delta/2^n,$$

and define

$$H_n = \bigcap_{i=1}^{n} E_i.$$

Since (H_n) is a contracting sequence of closed bounded sets, we can obtain the desired contradiction by showing that each set H_n is non-empty; for then the intersection of all the sets H_n would be non-empty even though the larger sets A_n have an empty intersection. For this

purpose, we make the following two observations: Firstly, for any n, if E is an elementary subset of $A_n \setminus E_n$, then since

$$m(E) + m(E_n) = m(E \cup E_n) \leqslant \alpha_n \quad \text{and} \quad m(E_n) > \alpha_n - \delta/2^n,$$

it follows that $m(E) < \delta/2^n$. Secondly, for any n, if E is an elementary subset of $A_n \setminus H_n$, then since

$$E = (E \setminus E_1) \cup (E \setminus E_2) \cup (E \setminus E_3) \cup \cdots \cup (E \setminus E_n)$$

and since $E \setminus E_i$ is an elementary subset of $A_i \setminus E_i$ for every $i = 1, 2, \dots n$, it follows that $m(E) < \delta$.

But for every n, because $\alpha_n > \delta$, the set A_n must have an elementary subset E such that $m(E) > \delta$, and so it follows that each set H_n is non-empty.

The Main Result. *Suppose* (f_n) *is a sequence of Riemann integrable functions on* $[a, b]$, *suppose* f *is a Riemann integrable function on* $[a, b]$, *that* $f_n \to f$ *pointwise on* $[a, b]$ *and that for some constant* $K > 0$, *we have* $|f_n| \leqslant K$ *for every* n. *Then we have*

$$\int_a^b f_n \to \int_a^b f.$$

Proof. There is no loss of generality in assuming that $f_n \geqslant 0$ for each n and that $f = 0$. Let $\varepsilon > 0$, and for each n, define

$$A_n = \left\{ x \in [a, b] \,|\, f_i(x) \geqslant \frac{\varepsilon}{2(b-a)} \quad \text{for at least one natural } i \geqslant n \right\}.$$

We now apply the lemma to (A_n) to choose a natural N such that whenever $n \geqslant N$, and E is an elementary subset of A_n, we have $m(E) < \varepsilon/2K$, and the proof will be complete when we have shown that whenever $n \geqslant N$, we have $\int_a^b f_n \leqslant \varepsilon$. Let $n \geqslant N$. Since the integral of a Riemann integrable function is the same as its lower integral, in order to show that $\int_a^b f_n \leqslant \varepsilon$, it is sufficient to show that whenever s is a step function and $0 \leqslant s \leqslant f_n$, we have $\int_a^b s \leqslant \varepsilon$. Let s be such a step function and define

$$E = \left\{ x \in [a, b] \,|\, s(x) \geqslant \frac{\varepsilon}{2(b-a)} \right\} \quad \text{and} \quad F = [a, b] \setminus E.$$

Then E and F are elementary sets, and since $E \subseteq A_n$, we have $m(E) < \varepsilon/2K$. Therefore

$$\int_a^b s = \int_E s + \int_F s \leqslant \int_E K + \int_F \frac{\varepsilon}{2(b-a)} \leqslant \int_E K + \int_a^b \frac{\varepsilon}{2(b-a)}$$

$$= Km(E) + \frac{\varepsilon}{2(b-a)}(b-a) < \varepsilon.$$

And that is all there is to it. Notice that while the above proof employs some of the notation and conveys some of the atmosphere of more advanced treatments of integration, it keeps well away from anything hard. Lebesgue measure is needed only for elementary sets; and all the measure is in this case is the sum of the lengths of the finitely many component intervals that make up an elementary set. The proof is accessible to students who have never seen countability and never seen infinite series. They don't even need the Heine Borel theorem if they know that a bounded sequence of real numbers must have a partial limit (cluster point) and that, consequently, a contracting sequence of non empty closed bounded sets must have a non empty intersection.

Incidentally, it is easy to adapt the above proof to show that even if it is not assumed that the limit function f is Riemann integrable, because $(f_n(x))$ is a Cauchy sequence for each x, the sequence of integrals $\int_a^b f_n$ must be a Cauchy sequence and must therefore converge. This may be used to give a revealing explanation of the inadequacy of the Riemann integral.

References

1. Tom M. Apostol, Mathematical Analysis, 2nd ed., Addison-Wesley, Reading, MA, 1974.
2. Robert G. Bartle and Donald R. Sherbert, Introduction to Real Analysis, Wiley, New York, 1982.
3. W. F. Eberlein, Notes on integration I: the underlying convergence theorem, Comm. Pure Appl. Math., Vol. X (1957) 357–360.
4. Walter Rudin, Principles of Mathematical Analysis, McGraw-Hill, New York, 1964.
5. ______, Real and Complex Analysis, McGraw-Hill, New York, 1966.
6. W. A. J. Luxemburg, Arzela's dominated convergence theorem for the Riemann integral, this MONTHLY, 78 (1971) 970–979.

EXPLAINING SIMPLE COMBINATORIAL ANSWERS

ROGER B. EGGLETON
Department of Mathematics, Statistics and Computer Science,
University of Newcastle, N.S.W. 2308, Australia

This note illustrates the principle: *If the answer to a problem turns out to be simple, there is probably a good explanation for it!* A simple answer should motivate us to try to derive that answer in a way which makes it obvious, or at least clarifies the underlying reason for its simplicity. Simplicity and clarity are of course subjective measures, but ones which are still useful. The practice of mathematics is an art as well as a science.

Consider combinatorics. Here it is recognized that simple answers are often satisfyingly explicable in terms of correspondences. This theme was taken up in [3], for example, from the viewpoint that counting the elements of a relatively unfamiliar set X can be satisfyingly achieved if we establish a correspondence between the elements of X and those of some relatively familiar set A. The correspondence constitutes the desired explanation. In this note we take up the theme from the viewpoint that explanations in terms of correspondences can also be achieved between two sets X and A of equally familiar structure. We illustrate this with several examples, most of which "explain" a well-known identity, and are therefore suitable for classroom use.

We shall use lower case symbols to denote natural numbers, including zero, and $I(n)$ will denote the set comprising the first n natural numbers (that is, the natural numbers less than n). The family of k-subsets of $I(n)$ will be denoted by $I(n, k)$. We regard the binomial coefficients as the cardinalities of such sets, by definition:

$$\binom{n}{k} := |I(n,k)|.$$

EXAMPLE 1 (*Symmetry of Pascal's Triangle*). Let $A := I(n, k)$ and $X := I(n, n-k)$. Pairing each k-subset of $I(n)$ with its complement gives a one-to-one correspondence $X \leftrightarrow A$. Hence $|X| = |A|$, so

$$\binom{n}{n-k} = \binom{n}{k}.$$

EXAMPLE 2 (*Pascal's Identity*, sometimes called *Vandermonde's Identity*). Let $A := I(n+1, k+1)$ and $X := X_0 \cup X_1$, where $X_0 := I(n, k)$ and $X_1 := I(n, k+1)$. Any $(k+1)$-subset of $I(n+1)$ either contains the element n or it does not. In the former case, pair it with the k-subset of $I(n)$ obtained by deleting the n, while in the latter case simply pair it with itself, now regarded as a $(k+1)$-subset of $I(n)$. This gives a one-to-one correspondence $X \leftrightarrow A$, since X_0 and X_1 are disjoint. Hence $|X| = |X_0| + |X_1| = |A|$, so

$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}.$$

EXAMPLE 3 (*Arithmetic Series Identity*). The sum of natural numbers up to n, inclusive, is $\frac{1}{2}n(n+1)$, which is a barely-disguised binomial coefficient. How can we explain the binomial

coefficient? The identity can be explained as two ways of counting the 2-subsets of $I(n+1)$, as follows. Let $A := I(n+1, 2)$ and let X_i be the subset of A comprising all 2-sets of the form $\{i, j\}$ with $i > j$. Then $\{i, j\} \leftrightarrow \{j\}$ gives a one-to-one correspondence $X_i \leftrightarrow I(i)$, whence $|X_i| = |I(i)| = i$. If $X := \bigcup_{i \leqslant n} X_i$, this union is disjoint and $X = A$, so $\sum_{i \leqslant n} |X_i| = |A|$, that is,

$$\sum_{i \leqslant n} i = \binom{n+1}{2}.$$

EXAMPLE 4 (*Diagonal Identities for Pascal's Triangle*). We generalize Example 3; let $A := I(n+1, k+1)$ and let X_i comprise all members of A with maximum element i. Each member of X_i can be paired with the k-subset of $I(i)$ resulting from deletion of i, which gives a one-to-one correspondence $X_i \leftrightarrow I(i, k)$, whence

$$|X_i| = |I(i, k)| = \binom{i}{k}.$$

Also $A = X := \bigcup_{i \leqslant n} X_i$, and the union is disjoint, so $|X| = |A|$ implies

$$\sum_{i \leqslant n} \binom{i}{k} = \binom{n+1}{k+1}.$$

EXAMPLE 5 (*Binomial Coefficient Identities of Second Order*). We now obtain some less familiar identities, by the line of reasoning developed in the previous examples. Let $m := r + s$, and let $A := I(n+1, m+1)$. Each member of A can be written in the form $\{a_0, a_1, \dots, a_m\}$, where $a_j < a_{j+1}$ for $j < m$. Let $X_i := I(i, r)$ and $Y_i := I(n-i, s)$. If $a_r = i$, we can pair the r smallest elements of $\{a_0, a_1, \dots, a_m\}$ with the corresponding r-subset of $I(i)$, and as the last s elements of $\{a_0, a_1, \dots, a_m\}$ are greater than i, we can subtract $i+1$ from each and so pair these elements with the resultant s-subset of $I(n-i)$. This pairing is

$$\{a_0, a_1, \dots, a_r = i, \dots, a_m\} \leftrightarrow (\{a_0, a_1, \dots, a_{r-1}\}, \{a_{r+1} - i - 1, \dots, a_m - i - 1\}),$$

which gives a one-to-one correspondence $A_i \leftrightarrow X_i \times Y_i$, where A_i is the subset of A comprising all $(m+1)$-sets with $a_r = i$. Hence

$$|A_i| = |X_i| \cdot |Y_i| = |I(i, r)| \cdot |I(n-i, s)| = \binom{i}{r}\binom{n-i}{s}.$$

Also $A = \bigcup_{i \leqslant n} A_i$, and the union is disjoint (indeed, its members with $i < r$ or $i > n - s$ are empty), whence the cardinalities yield

$$\sum_{i \leqslant n} \binom{i}{r}\binom{n-i}{s} = \binom{n+1}{m+1}, \quad \text{where } m := r + s.$$

We could derive this identity in a slightly more homogeneous form by putting $j := n - i$, so $Y_i := I(j, s)$ and the end result becomes

$$\sum_{i+j=n} \binom{i}{r}\binom{j}{s} = \binom{n+1}{m+1}, \quad \text{where } m := r + s,$$

the summation being over all ordered pairs (i, j) with $i + j = n$.

EXAMPLE 6 (*Binomial Coefficient Identities of Higher Order*). The method used in Example 5 readily yields third and higher order identities. For example, let $m := r + s + t$ and let $A := I(n+2, m+2)$. Arrange the elements of any $(m+2)$-subset of $I(n+2)$ in increasing order, and let A_{ij} comprise all those sets $\{a_0, a_1, \dots, a_{m+1}\}$ with $a_r = i$ and $a_{r+s+1} = i + j + 1$. Pair this set with the ordered triple comprising (1) the r-set consisting of the first r elements, (2) the s-set derived from the elements greater than i and less than $i + j + 1$, after subtracting $i + 1$ from each of them, and (3) the t-set derived from the last t elements, after subtracting $i + j + 2$

from each of them. Putting $k := n - i - j$ leads to

$$\sum_{i+j+k=n} \binom{i}{r}\binom{j}{s}\binom{k}{t} = \binom{n+2}{m+2}, \quad \text{where } m := r + s + t,$$

the summation being over all ordered triples (i, j, k) with $i + j + k = n$.

EXAMPLE 7 (*Diagonal Identities for Falling Factorials*). If objects other than subsets are used, we can obtain identities involving their counting numbers instead of binomial coefficients. We conclude with one such example. Let $S(n, k)$ denote the set of k-sequences of distinct terms, all from $I(n)$. The *falling factorial* $n_{(k)}$ can be defined as:

$$n_{(k)} := |S(n, k)|.$$

It is well known that $n_{(k)} = \prod_{i<k}(n - i)$, the case of the empty product taking the value 1 as usual. Let $A := S(n + 1, k + 1)$ and let A_i be the subset of A comprising sequences with maximum term i. If $\mathbf{a}$ is a $(k + 1)$-sequence in A_i, it can be paired with the ordered pair $(\mathbf{a}', r)$, where $\mathbf{a}'$ is the k-sequence obtained from $\mathbf{a}$ by deleting its maximum term, and r is the term number of that maximum term (so $a_r = i$). This gives a one-to-one correspondence $A_i \leftrightarrow X_i \times I(k + 1)$, where $X_i := S(i, k)$. Hence

$$|A_i| = |X_i| \cdot |I(k + 1)| = |S(i, k)| \cdot |I(k + 1)| = i_{(k)}(k + 1).$$

Also $A = \bigcup_{i \leqslant n} A_i$, and the union is disjoint, so the cardinalities yield

$$(k + 1) \sum_{i \leqslant n} i_{(k)} = (n + 1)_{(k+1)}.$$

Of course, this identity can be obtained mechanically from the one in Example 4 by multiplying the latter by $(k + 1)!$ but that does not achieve a direct "explanation" of the identity.

For elegant discussions of many of the basic identities of combinatorics, we recommend Berge's book [**1**] to the reader whose interest in combinatorics is just beginning. A more comprehensive treatment in a similar spirit is that of Comtet [**2**], while encyclopaedic treatments of combinatorial identities have been given by Riordan [**5**] and Gould [**4**].

References

1. C. Berge, Principles of Combinatorics (trans. John Sheehan), Academic Press, 1971.
2. Louis Comtet, Advanced Combinatorics, D. Reidel Publishing Company, 1974.
3. Romae J. Cormier and Roger B. Eggleton, Counting by correspondence, Math. Mag., 49 (1976) 181–186.
4. H. Gould, Combinatorial Identities, Morgantown Printing Company, 1972.
5. J. Riordan, Combinatorial Identities, Wiley, 1968.

---

169. MISCELLANEA

Bilateral Convolution

The right brain harbors the spark, the ignition:
intuition.
The left brain houses language, cognition,
the logician.
Two hemispheres in coalition—
anatomy of a mathematician.

Katherine O'Brien

PROBLEMS AND SOLUTIONS

For instructions about submitting ***proposed*** *problems for publication in this department see the inside front cover. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

Solutions *should be sent to the address given on the inside front cover.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

For instructions about submitting solutions of these Elementary Problems, which should be mailed by September 30, 1986, see the inside front cover. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgement).

E 3147. *Proposed by A. Wilansky, Lehigh University.*

Let (X, B, μ) be a measure space such that all singletons are measurable. Let $f(x) = \mu(\{x\})$. Must f be measurable?

E 3148. *Proposed by Rick Luttmann, Sonoma State University, Rohnert Park, CA.*

Let n distinct pairs of socks be put into the laundry. (It is assumed that each of the $2n$ socks has precisely one mate.) When the laundry is returned, the socks are drawn out one at a time; each is matched with its mate, if the mate has previously been drawn.

Find a formula for the expected number $E(k)$ of pairs formed after k socks have been drawn, $k = 1, 2, \ldots, 2n$.

E 3149. *Proposed by Louis Funar, University of Craiova, Romania.*

Let K be a plane convex set with area a, perimeter p and diameter d, and let $\lambda = \dfrac{a}{a + pd + \pi d^2}$. Prove that any finite family of congruent copies of K that covers area A must have a subfamily with pairwise disjoint interiors that covers area at least λA.

*Is this also true with $\lambda = \dfrac{a}{\pi d^2}$?

E 3150. *Proposed by George A. Tsintsifas, Thessaloniki, Greece.*

Let ABC be a triangle with sides a, b, c, and area F. It is well known that $a^2 + b^2 + c^2 \geqslant 4\sqrt{3}\,F$. If p, q, r are arbitrary positive real numbers, prove that

$$\frac{p}{q+r}a^2 + \frac{q}{r+p}b^2 + \frac{r}{p+q}c^2 \geqslant 2\sqrt{3}\,F.$$

E 3151. *Proposed by Péter Ivády, Institute for Economy and Organization, Budapest, Hungary.*

Let $x \geqslant 0$, $x \neq 1$, $\lambda \geqslant 1$ and $0 \leqslant \beta \leqslant 2$ be real numbers. Prove that

$$\left(\frac{x^\lambda - 1}{x - 1}\right)^\beta \leqslant \lambda\left(\frac{x^{\lambda\beta} - 1}{x^\beta - 1}\right).$$

E 3152. *Proposed by Leopoldo Nachbin, University of Rochester.*

Let S be a set of infinite sets. Consider the following partition property (PTP):

For every $X \in S$, there is an infinite subset $Y_X \subset X$ such that $Y_{X_1} \cap Y_{X_2} = \varnothing$ if $X_1, X_2 \in S$ are distinct.

Prove that:

(1) S has property (PTP) if it is countable;

(2) For every uncountable cardinal number N, there is some S whose power is N, but which fails to have property (PTP).

SOLUTIONS OF ELEMENTARY PROBLEMS

A Hypertrigonometric Inequality

E 2991 [1983, 212]. *Proposed by Z. F. Starc, Vrasc, Yugoslavia.*

Prove or disprove the following inequality:

$$\alpha \tanh(x) > \sin(\alpha x) \qquad (x > 0,\ \alpha > 1).$$

Composite solution. Expansions in Maclaurin series show that the inequality fails in a right neighborhood of the origin for $\alpha < \sqrt{2}$.

The inequality holds for sufficiently large x for any α in the given range. For it to fail, the left side minus the right side must have a minimum at which it is negative for positive x. The condition for a minimum, ($\operatorname{sech}^2(x) = \cos(\alpha x)$), and for the inequality to be violated, together lead to the inequality: $\alpha^2 < 2\cos^2(\alpha x/2)$, which cannot hold for $\alpha \geqslant \sqrt{2}$. The inequality therefore holds in the range ($x > 0$, $\alpha \geqslant \sqrt{2}$).

Thirty-eight correct solutions to the problem were received. Some solvers provided explicit counterexamples, some gave the Maclaurin expansion. O. P. Lossers (The Netherlands) submitted the proof for $\alpha \geqslant \sqrt{2}$ given above. O. G. Ruehr provided an integral representation of the difference between the sides of the inequality, with integrand explicitly nonnegative for $\alpha \geqslant \sqrt{2}$. Other proofs of the inequality for α at least $\sqrt{2}$ were provided by W. A. Newcomb, L. Kuipers (Switzerland), D. Caccia, and the Chico Problem Solving Group,

The Maximum Number of Zero Entries in a "Good" Matrix

E 2995 [1983, 334]. *Proposed by Miroslav D. Asič, London School of Economics.*

A square matrix of order n, $n \geqslant 2$, is said to be "good" if it is symmetric, invertible and all its entries are positive. What is the largest possible number of zero entries in the inverse of a "good" matrix?

Solution by A. A. Jagers, Technische Hogeschool Twente, Enschede, The Netherlands. Let z_n denote the largest possible number of zero entries in the inverse of a "good" matrix of order $n \geqslant 2$. Then $z_n = n^2 - 2n$.

Proof. The inner product of the jth row of a matrix A and the ith column of its inverse A^{-1} is equal to 0 for $i \neq j$, by definition of inverse. Hence, if A is "good", then each column of A^{-1} contains at least one negative entry and one positive entry, and so $z_n \leqslant n^2 - 2n$. This bound is met for the following "good" matrix of order n:

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 2 & 2 & \cdots & 2 \\ 1 & 2 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 1 & 2 & \cdots & 2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \\ 1 & 2 & 1 & 2 & & \end{bmatrix},$$

with tridiagonal inverse

$$A^{-1} = \begin{bmatrix} 2 & -1 & & & & \\ -1 & 0 & 1 & & & \\ & 1 & 0 & -1 & & \\ & & -1 & 0 & & \\ & & & & 0 & -s \\ & & & & -s & s \end{bmatrix},$$

where $s = (-1)^n$.

Also solved by J. P. Farges (Canada), P. M. Gibson, J. K. Haack, K. Kearnes, L. R. King, J. R. Kuttler, D. Lindsay, O. P. Lossers (The Netherlands), A. J. Schwenk, S. White, E. T. Wong, and the proposer.

A Theorem of Goldbach

E 2999 [1983, 335]. *Proposed by J. D. Shallit, University of California, Berkeley, and Karel Zikan, San Jose State University.*

Let S be the set of nontrivial integer kth powers, i.e.,

$$S = \{n^k | n \geqslant 2,\ k \geqslant 2\} = \{4, 8, 9, 16, 25, 27, 32, 36, \dots\}.$$

Evaluate $\Sigma(s-1)^{-1}$, the sum being extended over all members s of S.

Solution by University of South Alabama Problem Group. Let T be the set of all positive integers greater than 1 that are not in S. Then

$$\begin{aligned}\sum(s-1)^{-1} &= \sum_{k\geqslant 2}\sum_{a\in T}(a^k-1)^{-1} \\ &= \sum_{k\geqslant 2}\sum_{a\in T}\sum_{i\geqslant 1} a^{-ik} \\ &= \sum_{n\geqslant 2}\sum_{k\geqslant 2} n^{-k} \\ &= \sum_{n\geqslant 2}(n(n-1))^{-1} = 1.\end{aligned}$$

This result is an old theorem of Goldbach, as was noted by several solvers. Janous noted that the result was stated in 1729 in a letter from Goldbach to Daniel Bernoulli (see G. Klambauer, *Mathematical Analysis*, M. Dekker Inc., New York, 1975, p. 120, 38: or the recent biography *Christian Goldbach* 1690–1764, p. 164, written by A. P. Jushkevitch and Ju. Ch. Kopelevitch, Nauka, Moscow, 1983 (Russian)). Klamkin referenced Chrystal's *Algebra II* and Pei Yuan Wu noted that the evaluation of the generalized Goldbach-Euler series

$$\sum_{k=2}^{+\infty}\sum_{s=0}^{+\infty}\frac{1}{(ps+r)^k-1} = \frac{1}{p}\left[\psi\left(\frac{r}{p}\right) - \psi\left(\frac{r-1}{p}\right)\right] + \frac{1}{r(r-1)},$$

where $\psi(x) = \dfrac{\Gamma'(x+1)}{\Gamma(x+1)}$ denotes the digamma function and $r = p$ or $p+1$ can be found in Z. A. Melzak, *Companion to Discrete Mathematics: Mathematical Techniques and Various Applica-*

tions, John Wiley & Sons, New York, 1973, page 88.

Several solvers noted that the result can also be expressed as

$$\sum_{k=2}^{+\infty} (\zeta(k) - 1) = 1.$$

Also solved by D. M. Bloom, R. Breusch, Y. H. Fai (Australia), W. Janous (Austria), K. Kearnes, M. S. Klamkin (Canada), O. P. Lossers (The Netherlands), D. Moews, A. Schwenk, A. Tissier (France), J. B. Wilker (Canada), Pei Yuan Wu (Taiwan), and the proposers. Partially solved by P. S. Bruckman and M. Wyneken.

ADVANCED PROBLEMS

For instructions about submitting solutions of these Advanced Problems, which should be mailed by September 30, 1986, see the inside front cover. The solver's full post-office address should be on each sheet.

6518. *Proposed by Vladimir N. Akis, California State University at Los Angeles.*

Let S be a closed square of area 4. Denote by $\mathscr{C}$ the collection of all squares of area 1 contained in S whose centers are points of S, and whose sides are parallel to the sides of S.

Let $f: \mathscr{C} \to S$ be a choice function, where $f(A) \in A$ for each $A \in \mathscr{C}$. Suppose that f is continuous with respect to the Hausdorff metric on $\mathscr{C}$. Show that $f(A)$ is the center of S for some $A \in \mathscr{C}$.

6519. *Proposed by Ira Gessel, Brandeis University.*

Let

$$F(a, b, m, n) = \sum_{j=0}^{m} \binom{a+m+n-2j}{n-j}\binom{a+n}{j}\binom{b+m}{m-j},$$

where m and n are nonnegative integers. Show that $F(a, b, m, n) = F(b, a, n, m)$.

6520. *Proposed by Robert B. Israel, University of British Columbia.*

Show there exists a rational function f such that for every holomorphic function g in the unit disk D, g or $g - f$ has a zero in D. (See Problem 6437 [1983, 485; 1985, 365].)

SOLUTIONS OF ADVANCED PROBLEMS

A Joint Distribution Formula

6475 [1984, 588]. *Proposed by Ignacy Icchak Kotlarski, Oklahoma State University.*

Let $\{X(t): t \geqslant 0\}$ and $\{Y(t): t \geqslant 0\}$ be stochastic processes defined on a probability space $(\Omega, \mathscr{F}, \mathscr{P})$ with $X(t)$ and $Y(t)$ both having $N(0, t)$ distributions for all $t \geqslant 0$. Let Z be a random variable defined on $(\Omega, \mathscr{F}, \mathscr{P})$ with a $\chi^2(1)$ distribution such that $\{X(\cdot)\}, \{Y(\cdot)\}$ and Z are independent. Let $U = X(Z)$ [i.e., $U(\omega) = X(Z(\omega))(\omega)$, $\omega \in \Omega$] and $V = Y(Z)$. Find the joint distribution of (U, V).

Solution by K. B. Athreya, Department of Statistics, Iowa State University. Let $f_{U,V}(u, v)$ denote the joint probability density of (U, V). Since Z is independent of $X(\cdot)$ and $Y(\cdot)$, we obtain by the usual method of conditioning first on Z and then integrating that

$$f_{U,V}(u, v) = \frac{1}{(2\pi)^{3/2}} \int_0^\infty e^{-\frac{u^2+v^2}{2z}} e^{-z/2} z^{-3/2}\, dz$$

$$= \frac{1}{(2\pi)^{3/2}} \frac{1}{\sqrt{2}} \int_0^\infty e^{-\left(\frac{\lambda}{u}+u\right)} u^{-3/2}\, du,$$

where

$$\lambda = (u^2 + v^2)/4.$$

Upon differentiating the well-known integral identity

$$\int_0^\infty e^{-\left(\frac{\lambda}{u}+u\right)} u^{-1/2}\, du = \sqrt{\pi}\, e^{-2\sqrt{\lambda}}, \quad 0 < \lambda < \infty,$$

we find that

$$f_{U,V}(u, v) = \frac{e^{-R}}{2\pi R},$$

where

$$R = \sqrt{u^2 + v^2}, \quad -\infty < u, v < \infty.$$

Thus R has an exponential distribution with mean one, and given R the vector (U, V) has a uniform distribution on the circumference of the circle of radius R centered at the origin.

Allan Gutjahr points out that similar results can be obtained for cases where V has a chi-square distribution with $m > 1$ degrees of freedom. For example, when $m = 2$ the density (in Bessel function notation) is $K_0(R)/2\pi$.

K. B. Athreya observes that his method above does not require the independence of $X(\cdot)$ and $Y(\cdot)$. He adds the following generalization to k stochastic processes $\{X_i(t) : t \geqslant 0\}, 1 \leqslant i \leqslant k$, on a probability space (Ω, B, P). Assume that for each $t \geqslant 0$, the vector

$$X(t) = (X_1(t), \ldots, X_k(t))$$

is multivariate normal with mean 0 and covariance matrix of the form tS, where S is a fixed positive definite matrix. Let Z be a random variable also on (Ω, B, P) but independent of the $\{X_i(\cdot)\}$ for $1 \leqslant i \leqslant k$, and such that it has a gamma (α, p) distribution. Assume $n = (k + 1 - 2p)/2$ is a nonnegative integer. Then the joint probability density of

$$U \equiv (U_1, \ldots, U_k),$$

where $U_i = X_i(Z)$ is given by

$$f_U(u) = \frac{\alpha^{k/2}(-1)^n F^{(n)}\left(\frac{1}{2} u S^{-1} u^{\mathrm{T}} \alpha\right)}{(2\pi)^{(k-1)/2} \Gamma(p) (\det S)^{1/2}},$$

where $F^{(n)}(\cdot)$ is the nth derivative of $\exp(-2\sqrt{\lambda})$. This says that given $R = US^{-1}U^{\mathrm{T}}$, the vector U has the uniform distribution on the surface of the ellipsoid

$$\{u : uS^{-1}u^{\mathrm{T}} = R\}$$

and R has the probability density on $(0, \infty)$ given by

$$f_R(r) = \frac{\alpha^{k/2}(-1)^n F^{(n)}\left(\frac{r\alpha}{2}\right) c_k(r)}{(2\pi)^{(k-1)/2} \Gamma(p) (\det S)^{1/2}}$$

where $c_k(r)$ is the "surface area" of the ellipsoid

$$\{u : uSu^{\mathrm{T}} = r\}.$$

(Athreya also manages to remove the restriction that n be an integer, at the cost of introducing

certain families of Bessel functions.)

Also solved by Allan Gutjahr, Ellen Hertz, John Kieffer, Lajos Takács, Jovan Vukmirović (Yugoslavia), and the proposer.

A Delicate Remainder Sum

6476 [1984, 588]. *Proposed by Louis Funar, University of Craiova, Romania.*

For positive integers n and k, define $R(n,k)$ to be the remainder when n is divided by k. Let

$$S(n,k)=\sum_{i=1}^{k}R(n,i).$$

(a) Prove that $\lim_{n\to\infty}\frac{S(n,n)}{n^2}=1-\frac{\pi^2}{12}$.

(b)* Is it true that $\lim_{n\to\infty}\frac{S(2^n,n)}{n^2}=1-\frac{\pi^2}{12}$?

Partial solution by Robert Breusch, Amherst College. From the identities

$$\begin{aligned}S(m,n)&=mn-\sum_{i=1}^{n}[m/i]i\\&=mn-[m/n]\sum_{i=1}^{n}i-\sum_{m/n<k\leqslant m}\left(\sum_{i=1}^{[m/k]}i\right),\end{aligned}$$

it is not hard to deduce

$$S(m,n)=mn-[m/n]n(n+1)/2-(m^2/2)\sum_{k=[m/n]+1}^{\infty}k^{-2}+O(m\log n).$$

Part (a) easily follows by setting $m=n$ and summing Σk^{-2}. Part (b), if true, must be somewhat delicate, since it is false for all more slowly growing sequences $m=rn$ where

$$r=r(n)\to\infty\quad\text{and}\quad\frac{r\log n}{n}\to 0.$$

In this case,

$$S(rn,n)=rn^2/2-(r^2n^2/2)\sum_{k=r+1}^{\infty}k^{-2}+o(n^2)$$

and

$$\begin{aligned}\sum_{k=r+1}^{\infty}\frac{1}{k^2}-\frac{1}{r+1/2}&=\sum_{k=r+1}^{\infty}\frac{1}{k^2}-\sum_{k=r+1}^{\infty}\left(\frac{1}{k-1/2}-\frac{1}{k+1/2}\right)\\&=-\frac{1}{4}\sum_{k=r+1}^{\infty}\frac{1}{k^4-k^2/4}=O(r^{-3}),\end{aligned}$$

so

$$\lim_{n\to\infty}\frac{S(r(n)n,n)}{n^2}=\frac{1}{4}.$$

Several readers identified (a) either explicitly as problem 43, p. 56, of G. Pólya and G. Szegő, *Problems and Theorems in Analysis I*, Springer-Verlag, 1972, or implicitly in known estimates for averages of $\sigma(n)$, the sum of the divisors of n. The proposer has some numerical evidence that the

ratio of $S(n, n)$ to $S(2^n, n)$ is bounded above and below by finite positive constants.

Also partially solved by M. Deutsch, C. Georghiou (Greece), Werner P. Kohs, Andrzej Makowski (Poland), William A. Newcomb, Pei Yuan Wu (Taiwan), and the proposer.

Circles, Triangles, Squares, and the Golden Mean

6477 [1984, 588]. *Proposed by Louis Funar, University of Craiova, Romania.*

Let r be the radius of the incircle of an arbitrary triangle lying in the closed unit square. Prove or disprove that $r \leqslant (\sqrt{5} - 1)/4$.

Solution by L. Kuipers, Sierre, Switzerland. We begin by stating four basic properties of the inradius $r(\triangle ABC)$ of the triangle $T = ABC$; the vertices are assumed to be noncollinear.

(1) If $F(\triangle ABC)$ and $P(\triangle ABC)$ are the area and perimeter of T, then

$$r(\triangle ABC) = \frac{2F(\triangle ABC)}{P(\triangle ABC)}.$$

(2) If B, C, D are collinear and C is strictly between B and D, then

$$r(\triangle ABC) < r(\triangle ABD).$$

(3) Let $AB = AC$ and l be a line through A parallel to BC. Let A, D_1, D_2 be 3 points on l with D_1 between A and D_2. Then

$$r(\triangle D_2BC) < r(\triangle D_1BC).$$

(4) If $AB = AC$ and D is interior to $\triangle ABC$ and on the line joining A to the midpoint of BC, then

$$r(\triangle BDC) < r(\triangle ABC).$$

We now determine the possible locations in the square $QRSP$ of the vertices of a triangle $\triangle ABC$ with maximal inradius r. We may assume Q, R, S, P are $(0,0), (1,0), (1,1)$ and $(0,1)$, respectively. By (2) no vertex is an interior point. If two vertices, say A and B, are on the same side, then (2) shows they are the two vertices of that side, (3) shows that C is on their bisector, and then (4) that C is the midpoint of the opposite side. In this case the result follows from (1). Hence we may assume that no two vertices are on the same side: say no vertex is interior to QR and B, A, C are on QP, PS, and SR, respectively. Draw a line l through A parallel to BC and let M be the midpoint of BC. Choose N on l so MN is perpendicular to BC. Let K be the intersection of the line MN and the boundary of the square. By (3) and (4),

$$r(\triangle ABC) < r(\triangle NBC) < r(\triangle KBC)$$

(unless A already coincided with N or K, in which case the argument is shorter). Now $\triangle KBC$ has two vertices on the same side unless K is on PS. But then C and R (or else B and Q) must coincide: if not, translate the triangle parallel to the y-axis PQ until they do. This would give an extremal triangle with an interior vertex K.

Thus, we need only consider triangles $\triangle RBA$ with $AB = AR$, B on QP, and A on PS. If α is the angle between QR and BR, then

$$R = (0,1), \quad B = (0, \tan\alpha), \quad M = \left(\frac{1}{2}, \frac{1}{2}\tan\alpha\right)$$

and

$$A = \left(\frac{1}{2} + \tan\alpha - \frac{1}{2}\tan^2\alpha, 1\right).$$

It can be shown trigonometrically that r is maximal only when $\alpha = 0$. (Here, as with properties

(1)–(4), the editor omits the details.)

Aage Bondesen showed that for $p + q \leqslant 1$ the inradius of the triangle with vertices $(0,0), (p,1), (1,q)$ is less than that of the triangle with vertices $(0,0)$, $(1,0)$ and $(p+q,1)$. Jordi Dou investigated the case in which a rectangle replaces the square. Victor Pambuccian points out that this problem is automatically solvable by the Tarski algorithm (see, e.g., A. Seidenberg, *A new decision method for elementary algebra*, Ann. Math., 60 (1954) 366–369). On the other hand, the editor wishes to add that (a) problems at this level or higher have almost never been thus solved in real time, (b) there seems almost no hope that the algorithm will produce proofs with intuitive appeal (at least for humans), and (c) the algorithm cannot, even in principle, generalize results to n dimensions.

Also solved by Aage Bondesen (Denmark), Jordi Dou (Spain), and Esther Szekeres (Australia).

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER

Galois Theory. By Harold M. Edwards. Springer-Verlag, New York, 1984, xii + 152 pp.

PETER M. NEUMANN
The Queen's College, Oxford OX1 4AW, England

Galois was a Frenchman who made deep and important contributions to mathematics at the age of 17, who nevertheless failed the entrance examinations for the École Polytechnique, who was imprisoned at the age of 19 for political activities, and who died aged 20 in 1832, shot in an early morning duel. So romantic a figure has naturally attracted a great deal of posthumous attention. Biographers have done their best, but a person who dies young does not leave masses of documentary evidence behind and most of what has been written about Galois the man is more journalistic than historical; there are many theories but few facts. For different reasons Galois the mathematician is hardly better understood. His discoveries are so advanced that they are far beyond the mathematical competence of trained historians; and few mathematicians have the training in historical techniques or the taste for historical research that a study of his mathematics would require.

Galois theory, which is what the theory of equations was changed into by Galois' work, is one of the three great contributions that he made to mathematics. The other two were the theory of 'Galois imaginaries', which is, in essence, the same as the modern theory of finite fields, and the theory of groups, which emerged from his theory of equations and from Cauchy's papers of 1845 on the theory of substitutions. The theory of 'Galois imaginaries' was published when Galois was 18 years old in his paper 'Sur la théorie des nombres'. His work on theory of equations was first submitted to the Academy in 1829. It was either lost or withdrawn and a new paper was submitted in February 1830 for the *grand prix de mathématiques*. This second version was lost amongst the papers of Fourier. A third version entitled 'Mémoire sur les conditions de résolubilité des équations par radicaux' was submitted to the Academy in January 1831, was read by Poisson (and possibly also Lacroix), was rejected in July that year and returned to its author. This is the manuscript, the so-called *premier mémoire*, that was retrieved by his friend Auguste Chevalier after Galois' death, that Liouville first published in 1846, and which contains the original exposition of Galois theory.

At the beginning of the last century the central problem of the theory of equations was to find

a formula that would express a root of a polynomial equation

$$x^n + a_1x^{n-1} + \cdots + a_{n-2}x^2 + a_{n-1}x + a_n = 0$$

in terms of the coefficients $a_1, \ldots, a_{n-2}, a_{n-1}, a_n$. It was hoped, and indeed required, that such a formula would be algebraic: that is to say, it would involve no operations other than $+$, $-$, $\times$, $\div$, and $\sqrt[k]{}$ for various values of k. What was wanted were analogues of the formulae

$$x = -a_1 \quad \text{and} \quad x = \frac{1}{2}\left(-a_1 \pm \sqrt{a_1^2 - 4a_2}\right)$$

that solve the problem in case $n = 1$ or $n = 2$, respectively. For $n = 3$ and $n = 4$ such solutions had been known for over 250 years. In 1770–71 Lagrange appears to have come round to the idea that it might, just possibly, be the case that no such formula exists when $n \geqslant 5$. From 1799 to 1814 Paolo Ruffini, an Italian *savant*, published books and papers describing his proof of this impossibility, but his work was long, confused and confusing, and it had little influence on his contemporaries and successors. There were still many mathematicians who believed that the desired formulae should exist. Indeed both Abel in 1823 and Galois in 1828 believed that they had solved the quintic equation before they found their own errors and went on to elucidate the real state of affairs.

What Abel did was to prove in 1824 that there is no formula of the required kind for equations of degree 5. Later he amplified his proof and extended it to cover explicitly equations of any degree $n \geqslant 5$. It is a common myth that Abel did this by proving the simplicity of the alternating groups A_n for $n \geqslant 5$. He did not: the idea of a group did not yet exist and Abel did not invent it. There was no group theory needed for Abel's proof. What he used was a certain fact about the possibilities for the number of different functions that can be obtained from a given function of n variables by permuting them amongst themselves; and this fact, although it has a natural and obvious group-theoretical setting, had been proved by Cauchy, in a paper published in 1815, by simple calculations with permutations.

It was left to Galois, almost certainly in 1829, to discover the need for groups in the theory of equations (though Galois himself was later aware that Abel, who died in 1829, had probably been thinking along similar lines about a year earlier). In the *premier mémoire* he treats equations with numerical coefficients and addresses a much more delicate question than Abel had answered. The question is, can a solution of such an equation be obtained by computing with known quantities, using addition, subtraction, multiplication, division and extraction of roots as the only allowable arithmetical operations; that is to say, does there exist a solution in terms of radicals? The answer is sometimes yes (as in the case of the equation $x^5 - 2 = 0$), and sometimes no (if by "known quantities" we mean rational numbers, then the equation $x^5 - 4x + 2 = 0$ is an example). Galois discovered that there is a group naturally attached to each equation, and he was able to analyse how that group changed when the domain of known quantities was extended and to produce a necessary and sufficient condition, expressed in group-theoretic terms, for the solubility of the given equation by radicals.

The *premier mémoire* has never been an easy paper to understand. It defeated Poisson in 1831. And the early exegetes, Liouville, Betti, Serret,..., had, in effect, to re-work the theory for themselves, though of course following (and acknowledging) the very explicit indications that Galois gives in his sequence of lemmas and propositions. The difficulty lies in his exposition. Globally he organises the theory very beautifully and straightforwardly, but locally his explanations show his extraordinary impatience. The whole paper is really no more than a sketch. The idea of a group, for example, is one of Galois' great innovations, but one learns what he has in mind only by reading the proof of his *Proposition I* and working backwards. It is true that at the end of the collection of definitions with which the published versions of the *premier mémoire* begin there is a brief passage setting out his terminology for permutations, substitutions and groups. But the explanation is exiguous in the extreme. Moreover, this is a late insertion that was

added on that dreadful night before the fatal duel. A year earlier, when the manuscript had been read by the Academy referees, it had contained no explanation of what a group was supposed to be.

The unfavourable report by those referees, Poisson and Lacroix (though it seems doubtful if Lacroix had much to do with it), may be read in the published *Procès-verbaux* of the *Académie des Sciences*, *Séance du* 4 *Juillet* 1831. Poisson and Lacroix criticise the work on two very solid grounds. One of these is, in general terms, the criticism that I have made above, expressed like this:

> Quoiqu'il en soit, nous avons fait tous nos efforts pour comprendre la démonstration de M. Galois. Ses raisonnements ne sont ni assez clairs, ni assez développés pour que nous ayons pu juger de leur exactitude,... .
>
> [Be that as it may, we have made every effort to understand Mr Galois' proof. His reasoning is neither clear enough nor far enough developed for us to have been able to judge its correctness,... .]

The referees' other criticism is quite different. On the title page of his manuscript Galois had written

> On trouvera ici une <u>condition</u> générale à laquelle <u>satisfait toute équation soluble par radicaux</u>, et qui réciproquement assure leur résolubilité. <u>On en fait l'application seulement aux équations</u> dont le degré est un nombre premier. Voici le théorème donné par notre Analyse:
>
> Pour qu'une équation de degré premier, qui n'a pas de diviseurs commensurables, soit soluble par radicaux, il <u>faut</u> et il <u>suffit</u> que toutes les racines soient des fonctions rationnelles de deux quelconques d'entre elles.
>
> [Here will be found a general <u>condition</u> that is <u>satisfied by all equations soluble by radicals</u>, and which conversely ensures their solubility. <u>Just one application is given, to equations of</u> prime degree. Here is the theorem given by our analysis:
>
> In order that an equation of prime degree, which has no rational divisors, shall be soluble by radicals it is <u>necessary</u> and <u>sufficient</u> that all the roots should be rational functions of any two of them.]

Poisson and Lacroix reacted as follows:

> ... en admettant comme vraie la proposition de M. Galois, on n'en serait guère plus avancé pour savoir si une équation donnée dont le degré est un nombre premier est résolue ou non par des radicaux, puisqu'il faudrait d'abord s'assurer si cette équation est irréductible, et ensuite si l'une de ses racines peut s'exprimer en fonction rationnelle de deux autres. La condition de résolubilité, si elle existe, devrait être un caractère extérieur que l'on pût vérifier à l'inspection des coefficients d'une équation donnée, ou, tout au plus, en résolvant d'autres équations d'un degré moins élevé que celui de la proposée.
>
> [... accepting Mr Galois' proposition as true, one is hardly further forward towards knowing if a given equation whose degree is a prime number is soluble by radicals or not, because one must first decide whether this equation is irreducible and then whether one of its roots may be expressed as a rational function of two others. The condition for solubility, if it exists, should be an external characteristic that one can verify by inspection of the coefficients of the given equation, or at least, by solving other equations of lower degree than the one proposed.]

They concluded their report by advising that, since the author said that his main proposition was part of a more general theory, and since it was often the case that a complete theory was easier to

understand than isolated parts of it, one should wait until the whole of the author's work was available before forming a final opinion; but as it then was, the part that had been submitted was not in a suitable state that they could recommend it for the Academy's approval.

With hindsight one may feel that this report was wrong. But I cannot think so: it seems to me to be a model of good refereeing. Can any of us be sure that in an analogous situation today we would react differently? I doubt it: it is an admirable report, sympathetic but firm. All that is wrong with it is that it deals with the work of an exceptionally brilliant and awkward man. Galois had no research supervisor who might have shown him how his discoveries should be properly written up. Besides, Galois was not a man who took advice easily. Another young mathematician might have taken the criticism to heart, re-written his work, published it and become famous. Galois took offence, returned to political agitation, died young and became famous.

After Galois' principal works were published by Liouville in 1846 his theory as it applied to polynomial equations was rapidly understood (understanding of the theory of groups beyond the little that had direct application to the study of equations grew rather more slowly), and it has been taught and learned with enthusiasm and pleasure ever since. Over the years the outward form of Galois theory has changed enormously. At the hands of Dedekind, Emmy Noether, Emil Artin and others the subject has moved away from its very concrete origins in the theory of equations and has metamorphosed into that part of abstract algebra that deals with fields and their automorphism groups. Today's student will find that the *premier mémoire* looks very different from Galois theory as it is to be found in modern texts such as Herstein's *Topics in algebra* (Chapter 5) or Stewart's *Galois theory*.

The book *Galois theory* by Harold Edwards is quite different from these. Like them it is an excellent textbook suitable for advanced undergraduates, but it takes the student back to those first few decades of the nineteenth century and returns to Galois' original conception of the subject. Notation is very similar to that of Lagrange and Galois; division of the text into short sections numbered consecutively is true to the style of that time. The first thirty sections contain an account of what had been done before Galois on cubic, quartic and cyclotomic equations and on symmetric polynomials, with particular reference to the work of Newton, Lagrange and Gauss. The remainder of the book is, in effect, a very much expanded version of the *premier mémoire* put into context and carefully explained. Edwards develops the theory in the form that Galois wrote it except that every lemma and theorem is properly proved, every i is dotted and every t is crossed. Where Galois was impatient and obscure Edwards is extraordinarily patient and clear. Even the question of how Galois groups may be calculated and how in principle, if not in practice, one may decide whether or not a given equation is soluble by radicals, is carefully and extensively treated (by methods due to Kronecker fifty years after Galois' death). This is, at last, a satisfactory and decisive response to the referees' two criticisms of Galois' work.

In the preface Edwards writes that he wanted to explain the theory "in terms close enough to Galois' own to make his memoir accessible to the reader". It is an admirable intention and one in which he will, I believe, be found to have fully succeeded. Nevertheless, it is perhaps a slightly dangerous one. Galois' own exposition is so sketchy that one might complete his arguments in several different ways and one can never be sure quite how much importance he himself attached to this point or that. It is no surprise therefore that I find myself cheerfully disagreeing with Harold Edwards on some details. Here are two examples.

(1) I see no historical justificaion for singling out the concept of "Galois resolvent" and giving it that name. All that Galois uses is the existence of a rational function t of the roots $a, b, c, \ldots$ of his polynomial equation $f(x) = 0$ which has the property that each root can be expressed as a function of t. (In modern terms t is a generator of the splitting field of $f(x) = 0$ and its existence is guaranteed by the "Theorem of the primitive element." This is what Edwards [p. 35] calls a "resolvent" of the equation $f(x) = 0$; what he calls a "Galois resolvent" is a very special type of "resolvent".) Galois is very offhand in his proof of the existence of such functions; in his paper 'Sur la théorie des nombres' that was published in 1830 he dismisses the existence as being clear.

Furthermore, Galois himself acknowledges that the existence of such functions t was known to Abel before him.

(2) The proof of *Lemme III* is a splendidly controversial matter. Poisson was unable to understand it and made a note on the manuscript to say so, but he accepted that the lemma was true by a result of Lagrange. Galois, incensed, appended "On jugera" to Poisson's note. Edwards feels that Galois was right and he gives a line of argument that undoubtedly completes the proof. But to do this he has to read very much more than is there into what Galois actually wrote, and I find his justification rather far-fetched. On balance I side with Poisson: it was up to Galois to be both clear and correct, whereas what he wrote is far too easily misunderstood.

I have other small criticisms of the book. For example, although it is primarily a contribution to mathematical exposition, not to the history of mathematics, I would have liked to see a paragraph or two about Abel's contributions and his influence (or lack of it) on Galois. Then again, the explanation of what is meant by solubility of cyclotomic equations by radicals is not entirely happy: elsewhere 'solution by radicals' involves using roots of equations $x^p - k = 0$, so why is a root of the equation $x^{p-1} + x^{p-2} + \cdots + x + 1 = 0$ not immediately acceptable as a radical in virtue of the fact that it is a root of $x^p - 1 = 0$? But all that these criticisms prove is that the author is right when he advises his students to 'Read the masters.' The reader must form his own judgment after reading what Galois and Harold Edwards themselves have written. That is one of the many points on which I am in complete agreement with him.

At the end of his famous testamentary letter, written on the night before the duel, Galois commends his manuscripts to Chevalier's care and writes

il se trouvera, j'espère, des gens qui trouveront leur profit à déchiffrer tout ce gâchis.

[there will, I hope, be people who will find it profitable to decipher all this mess.]

With his latest book Harold Edwards joins the select band of these *gens*. He has added another significant item to the new *genre* of mathematical publication that he created with his two earlier books *The Riemann Zeta Function* and *Fermat's Last Theorem*. Just as Galois' paper '... résolubilité des équations par radicaux' is very aptly named, so *Galois theory* has an unusually accurate title: this is not only a splendid textbook of that subject, but also an excellent contribution to the study of Galois the mathematician.

Winning Ways for Your Mathematical Plays, Volumes I & II. By Elwyn R. Berlekamp, John H. Conway, and Richard K. Guy. Academic Press, New York, 1982. Volume I, xxxi + 426 pp.; Volume II, xxxi + 424 pp.

ROBERT CONNELLY
Department of Mathematics, Cornell University, Ithaca, New York 14853

"It's only a game, like dying is only death."—Tom Paxton

Imagine the following offhand conversation between two erudite mathematicians, Right and Left, at a prestigious institution of higher learning, as they pick up their mail.

Left: "Wow! Look at this! A book that shows you how to win at Dots-and-Boxes!"

Right (somewhat bored): "Splendid. Now you can win against your seven-year-old."

Left (unperturbed, but defiant): "Dots is a subtle game".

Right (sarcastically): "Umm. I'll bet. You should write a paper on it."

Left: "Look. The book is loaded with fascinating mathematical games."

Right: "I don't like games. I give at the office."

Left: "This *is* the office."

Mathematicians (or perhaps mathematics itself) have a strong love-hate relationship with mathematical games. Witness "game theory". Axioms are the rules of the mathematics game. Theorems are simply statements, "I win."

On the other hand, it is embarrassing to be caught playing a game (or mathematics) that is "trivial" or "too easy". It is almost cheating. And even if your game is "difficult", it helps if you can justify it with something about importance, relevance, etc. Being "just a game", in some eyes, makes it almost automatically "not serious mathematics", "child's play".

Mathematicians can be classified as either games players or not. There seems to be a reasonably large proportion of Chess players, Go players, etc., among mathematicians. For them games and mathematics are too compelling to deny.

On the other hand, many feel as Right above that "serious" games are too much like mathematics itself and "do not provide adequate diversion to their mathematical interests." If you believe that, they can explain away uncounted other onerous character flaws as well.

Winning Ways is not for the games hater or the person who wants a tidy theory neatly worked out in the greatest possible generality. Conway's previous book, *On Numbers and Games*, affectionately called ONAG, is a bit more in this mathematical style. *Winning Ways* is chaotic, rich with more examples than you can digest, full of diagrams, cartoon pictures, tables, in-jokes, silly jokes, silly puns, mnemonic devices, who knows what other devices, and it has more information and game winning strategies than you would ever want to know. In short, it is more fun than the Three Stooges, albeit more intellectual. For example, the following figure copied from WWI (*Winning Ways*, volume I), page 121, gives some of the flavor of the book.

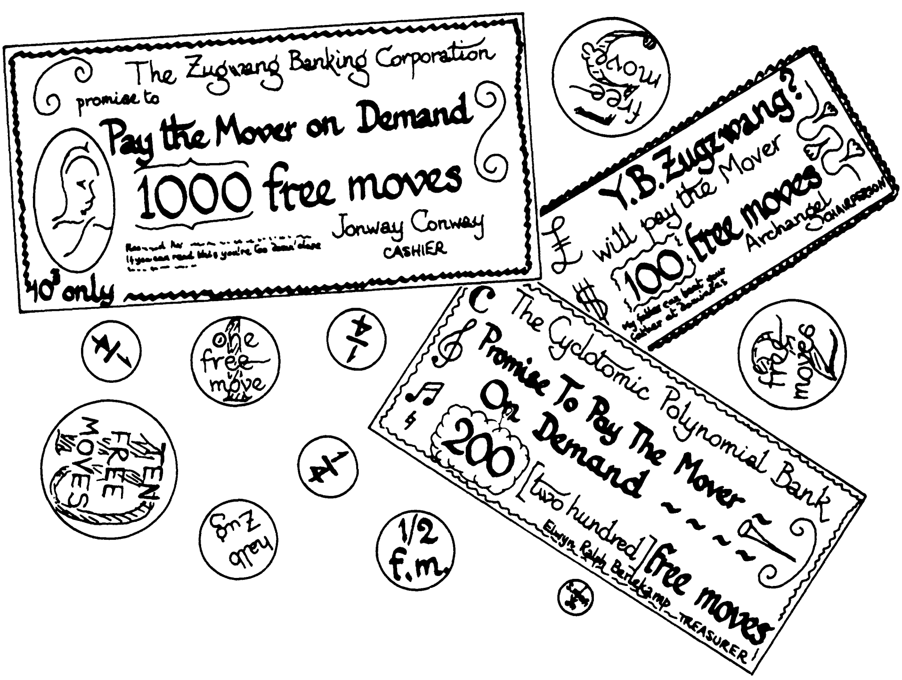


Figure 5. Some Cheques Ready for Cashing.

Zugzwang is a term, often used in Chess, for a position where a player does not want to move. The small print displays some bumper sticker mentality. $\uparrow$ is a game called up.

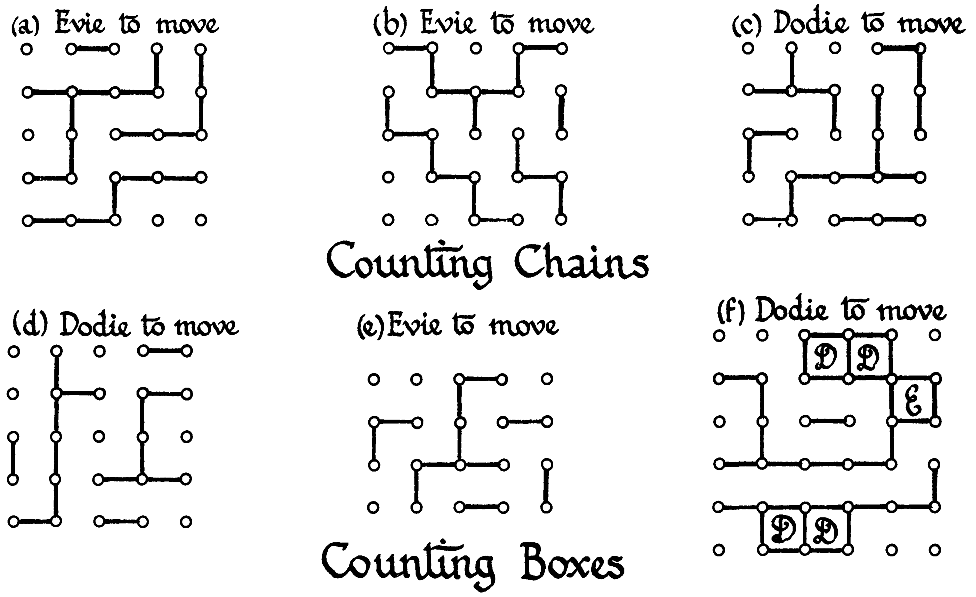
The book is unique. It can be read as a source of interesting combinatorial games (never mind a winning strategy), especially the second volume; one can just look at the funny cartoons; or it can provide much more. In the preface we have the following gratefully appreciated self-review:

> "We can supply the reviewer, faced with the task of ploughing through nearly a thousand information-packed pages, with some pithy criticisms by indicating the horns of the polylemma the book finds itself on. It is not an encyclopedia. It is encyclopedic, but there are still too many games missing for it to claim to be complete. It is not a book on recreational mathematics because there's too much serious mathematics in it. On the other hand, for us, as for our predecessors Rouse Ball, Dudeney, Martin Gardner, Kraitchik, Sam Loyd, Lucas, Tom O'Beirne and Fred. Schuh, mathematics itself is a recreation. It is not an undergraduate text, since the exercises are not set out in an orderly fashion, with the easy ones at the beginning. They are there though, and with the hundred and sixty-three mistakes we've left in, provide plenty of opportunity for reader participation. So don't just stand back and admire it, work of art though it is. It is not a graduate text, since it's too expensive and contains far more than any graduate student can be expected to learn. But it does carry you to the frontiers of research in combinatorial game theory and the many unsolved problems will stimulate further discoveries."

Winning Ways is, however, far weightier than its predecessors, both by mass and mathematics. The following are two of the many games discussed:

Nim: Two players start with three piles of beans, the piles having three beans, four beans, and five beans respectively. Each player in his turn must take some number of beans, but at least one bean, from one of the remaining piles. The player to remove the last bean wins. The idea behind the winning strategy is simple, and considerable generalizations of it are at the heart of much of this kind of game theory.

Dot-and-Boxes: This is the child's game where two players start with a rectangular array of dots and take turns joining a pair of horizontal or vertical dots. When a player completes the fourth side of a unit square, he puts his initial in that box and then must draw another line. The player at the end with the greater number of boxes wins.



Apparently this game can be played at about seven levels of sophistication and expert play involves a knowledge of much of the contents of WWI, which involves, among many other things, Sprague-Grundy Theory, which relates all impartial games to Nim.

The figure on page 413 shows some Dots-and-Boxes problems for advanced players, taken from WWII, page 536.

In addition to Nim-type games, much of WWI involves a development of Conway's idea of regarding number systems as particularized forms of games, as described in ONAG, but continued and expanded. Suppose two players Left and Right (no relation to those above) take turns playing a game, where at a position P the options to Left are $a, b, c, \dots,$ and the options to Right are $d, e, f, \dots$. Conway denotes P as $\{a, b, c, \dots / d, e, f, \dots\}$, reminiscent of Dedekind cuts. In fact, some games can be regarded as the real numbers. It is surprising how this simple idea leads to so many useful insights in this theory.

In WWII there is a discussion of Rubik's Cube, Conway's Life (neither a game nor a puzzle, but never mind), sliding block puzzles, Fox and Geese, Spots and Sprouts, String and Wire puzzles, and much, much more. Mercifully, Chess and Go are not discussed. (Nevertheless, the end-game in Go might benefit from some analysis as is in WWI.)

Needless to say, this book is winningly overwhelming.

LETTERS TO THE EDITOR

For instructions about submitting letters for publication in this department see the inside front cover.

Editor:

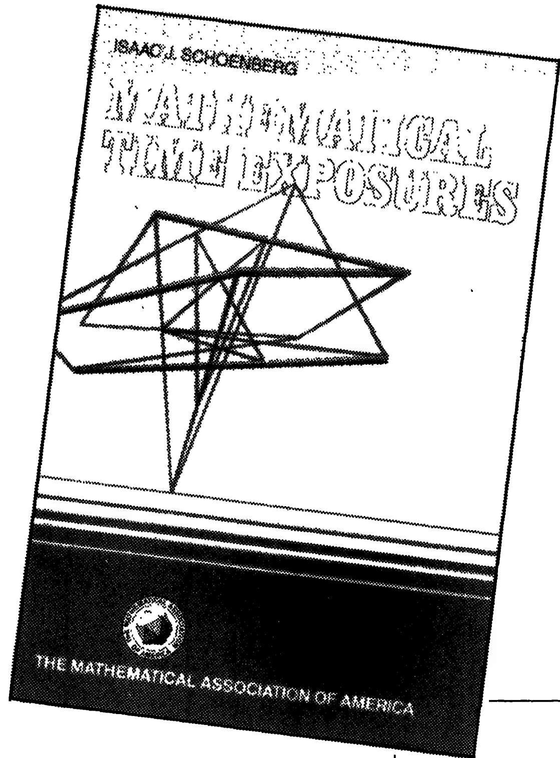
I am sure that by now you will have received several letters on the subject of Arthur Richert's note, "A Non-Simpsonian use of parabolas in numerical integration" (this MONTHLY, 92 (1985) 425–426). The statement that "the error bound is an improvement over Simpson's rule by a factor of almost eleven" is quite misleading. In fact, as the author himself points out, the Taylor polynomial method requires twice as many function evaluations for a given n as does Simpson's rule. Thus, it would be fairer to poor old Simpson to use $2n$ subdivisions in the error estimate, and this will reduce the Simpson rule error by a factor of approximately 16. Viewed in this light, Simpson's rule is the better rule, a fact which is easily borne out by experiment; e.g., taking $n = 20$ in Richert's formula and $n = 40$ in Simpson's rule, one gets approximations to π of

$$3.14159\ 26536\ 07233 \text{ and } 3.14159\ 26535\ 80105$$

using the usual integral for arctan(1) and (IEEE) 64 bit floating point arithmetic ($\pi = 3.14159\ 26535\ 89793\dots$). Here, the error of the Richert method is almost twice the error of Simpson's rule. The author's method is even more costly in this case, since the second derivative $f''(x)$ of $1/(1 + x^2)$ is computationally more complex than $f(x)$.

Paul Putter
Department of Mathematics and Computer Science
Drexel University
Philadelphia, PA 19104

Available in Hardcover
and Paper Editions



Mathematical Time Exposures

by Issac J. Schoenberg
279 pages
Hardcover Edition: List: $34.50
MAA Member: $26.00
Paper Edition List: $21.00
MAA Member: $15.50

Mathematical Time Exposures was inspired by Hugo Steinhaus' admirable book, *Mathematical Snapshots,* published in 1938. The title, *Mathematical Time Exposures* was also suggested by photography, but Schoenberg's pace is much more leisurely than Steinhaus'. Schoenberg spends more time on fewer subjects—the ''snapshots'' become ''time exposures.'' The subject of at least two of the chapters actually antedate the invention of the daguerreotype.

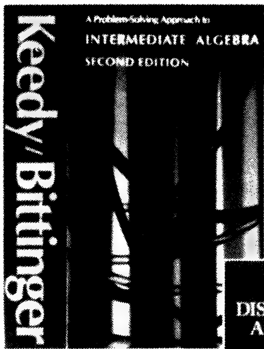
The author manages to bring together concepts from geometry, number theory,.algebra, and analysis, frequently mixing them together in the same chapter. The arts are not neglected. Discussions on the tuning of keyboard instruments, the guitar, and the vibrations of strings are discussed, as well as the suggestion of rectilinear models for outdoor sculpture.

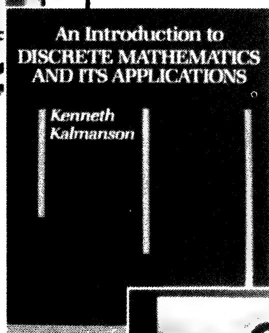
Many of the chapters in this book may be read independently, and at least half of them with only a knowledge of precalculus mathematics.

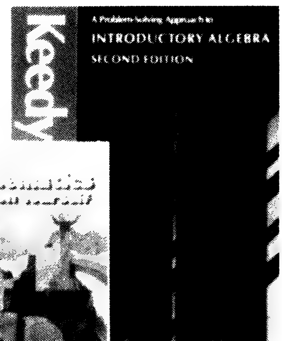
ORDER FROM:
The Mathematical Association of America
1529 Eioghteenth St, NW
Washington, DC 20036

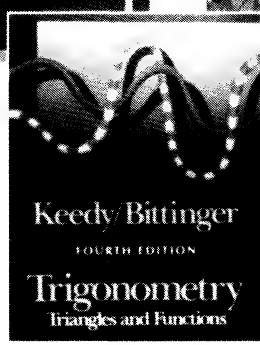
Brand New in '86

From Addison-Wesley, The Leading Publisher of Mathematics!









Coppins and Umberger — *Applied Finite Mathematics*

College Mathematics

DeGroot — *Probability and Statistics, Second Edition*

Falstein — *Basic Mathematics: You Can Count on Yourself*

Fuller — *Analytic Geometry, Sixth Edition*

Kalmanson — *An Introduction to Discrete Mathematics and its Applications*

Keedy-Bittinger — *A Problem Solving Approach to Introductory Algebra, Second Edition*

A Problem Solving Approach to Intermediate Algebra, Second Edition

Trigonometry: Triangles and Functions, Fourth Edition

College Algebra: A Functions Approach, Fourth Edition

Algebra and Trigonometry: A Functions Approach, Fourth Edition

Lance and Hinton — *Elementary Mathematics for Computing*

McHale, Christenson, and Roberts — *Introductory Algebra: Programmed, Second Edition*

Intermediate Algebra: Programmed, Second Edition

Moses — *Think and Explain with Statistics*

Oberle — *Calculus and the Computer*

Addison-Wesley Publishing Company

The Leader in Mathematics

And from our best-selling backlist...

Angel and Porter — *A Survey of Mathematics with Applications*

Fraleigh — *Calculus with Analytic Geometry, Second Edition*

Thomas and Finney — *Calculus and Analytic Geometry, Sixth Edition (Parts One and Two)*

To order, contact:

Addison-Wesley Publishing Company
1 Jacob Way
Reading, Massachusetts 01867
(617) 944-8660

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, DC 20036

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 93, Number 6 June-July 1986

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. (This is necessary for indexing purposes.) Send all main articles, letters to the editor, and editorial correspondence to the editor-elect: H.S. WILF, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Contributions to the other four special departments (Notes, Teaching of Mathematics, Unsolved Problems, and Problems and Solutions) should be sent to the appropriate one of the associate editors, as follows:

Notes: CAROL G. CRAWFORD, Department of Mathematics, U.S. Naval Academy, Annapolis, MD 21402, *or* RICHARD LIBERA, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716, *or* ANITA E. SOLOW, Department of Mathematics, Grinnell College, Grinnell, IA 50112.

Teaching of Mathematics: JOAN P. HUTCHINSON *or* STAN WAGON, Department of Mathematics, Smith College, Northampton, MA 01063.

Unsolved Problems: RICHARD GUY, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.

Proposed problems and solutions, both elementary and advanced: PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801.

Reprint Permission: A. B. WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Back Issues: P. and H. BLISS Co., Middletown, CT 06457.

Microfilm Editions: University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

Editorial Assistants: Fanny Hailpern, Mary P. Jackson.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT, ten times a year: January, February, March, April, May, June-July, August-September, October, November and December.

The annual subscription price for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is $24 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $24. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $70 per year.

Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

PRINTED IN THE UNITED STATES OF AMERICA

AN ALGORITHM FOR SOLVING POLYNOMIAL EQUATIONS

SYLVAN BURGSTAHLER

Department of Mathematical Sciences, University of Minnesota-Duluth, Duluth, Minnesota, 55812

1. Introduction. This paper discusses an algorithm for finding roots of a class of algebraic equations that includes polynomial equations. Like Newton's Rule, it is a second-order rule when used to find isolated roots. It also resembles Newton's Rule in degenerating to first-order in the case of multiple roots. However, algorithms of the same order of accuracy can behave quite differently! While this one involves more computation at each iterative step than Newton's Rule and is less versatile as well—not applying, for example, to transcendental equations—it usually determines roots of polynomial equations to a given accuracy in fewer steps and hence is often faster overall. In particular, it significantly outperforms Newton's Rule whenever the following conditions obtain:

* The modulus of the root is either very large or very small.
* The equation being solved has one or more zero coefficients either immediately following the term of highest degree or immediately preceding the constant term.

(1)

* The initial estimate is far from all of the roots.
* The equation has relatively few non-zero coefficients between the term of highest degree and the constant term and those few are small compared to the first and last terms.

If two or more of these conditions are met, the advantage over Newton's Rule can be spectacular. For example, if one initiates the solution of the equation

$$x^8 + x^4 + ix^3 + (2 - 11i)x^2 + 8x + 10^{10} = 0 \tag{2}$$

at (100, 200), the new algorithm finds the root near (6.8, 16.4) to more than twelve place accuracy in three iterative steps whereas Newton's Rule doesn't achieve *one* place accuracy until the 21st iteration!

The *modus operandi* of this algorithm is also unusual. Instead of fixing the equation and seeking improved root estimates as other algorithms do, this algorithm uses the preceding root estimate to modify the original equation. The resulting 'surrogate' equation is then solved for the root estimate needed to continue the process. Since polynomial equations are notoriously ill-conditioned—that is, minute changes in coefficients can have a tremendous effect on the roots [**1**]—such a strategy might seem doomed to failure. Seeing why it succeeds in spite of this gloomy prognosis leads to a deeper understanding of the concept of ill-conditioning as it applies to polynomial equations.

2. Some Algebraic Preliminaries. In its simplest form, the new algorithm creates surrogate equations by eliminating the interior terms of the original polynomial equation, i.e., all of the coefficients between the term of highest degree and the constant term. This process uses a result that will be given in more general form in Section 6 but, for now, the following form is sufficient:

LEMMA 1. *If* $|x - R| \ll |R|$, *then* $\left(\dfrac{x}{R}\right)^p \approx \dfrac{p}{n}\left(\dfrac{x}{R}\right)^n + \dfrac{(n-p)}{n}$.

Sylvan Burgstahler: I earned my Ph.D. at the University of Minnesota, Minneapolis, for work done under P. C. Rosenbloom and J. Serrin on the conduction of heat in composite solids. My outside interests are spread across everything from collecting (hand tools, rifle and pistol cartridges, and historic newspapers among other things) through reading (Mencken's *On Politics* being a representative favorite) and travel (most of the Civil War and World War I battlefields) to gardening (sometimes even producing ripe tomatoes despite an inhospitable climate). In short, knowing something about practically everything and a great deal about practically nothing, I am admirably equipped for my present roles as a grandfather and as a player of Trivial Pursuits.

Proof. If one defines $f(x)$ to be the difference between the left and right side of the conclusion, the (finite) Taylor's series expansion of $f(x)$ in terms of $(x - R)$ will be found to have neither zeroth nor first order terms. The coefficient of $(x - R)^2$ is $p(p - n)/(2R^2)$, from which the result follows. ∎

We are not yet ready to describe the new algorithm for general polynomial equations but for sparse polynomial equations—that is, those with relatively few interior terms—Lemma 1 is sufficient. To show its use and to illustrate the approach the algorithm takes, consider the following example:

$$x^6 - 9x^4 + 12x - 1639924 = 0. \tag{3}$$

This equation has a root at $x = 11$ that we will estimate using $x_0 = R = 10$.

To obtain the surrogate equation for (3) use Lemma 1 with $p = 4$ and $n = 6$ to eliminate the term $(-9x^4)$, and again with $p = 1$ and $n = 6$ to eliminate the term $(12x)$. The elimination of $(-9x^4)$ contributes $(-.06)$ to the sixth degree term and (-30000) to the constant term, whereas the elimination of $(12x)$ contributes 0.00002 to the sixth degree term and 100 to the constant term.

Summing these contributions and renaming the variable x_1 one obtains the surrogate equation

$$.94002x_1^6 - 1669824 = 0. \tag{4}$$

Equation (4) has six roots located on a circle in the complex plane. Picking the root closest to x_0, we find $x_1 \approx 11.004972$. Even though only the first approximation, this estimate is already impressively accurate. By way of contrast, Newton's Rule with $x_0 = 10$ would have produced $x_1 \approx 11.29395$, a result having almost 600 times greater error.

To iterate, return to (3) and use Lemma 1 with $R = x_1$ to re-eliminate the terms $(-9x^4)$ and $(12x)$. The result is a new surrogate equation of the same form as (4). Picking the root closest to x_1 yields $x_2 \approx 11.000000117$ which is stunningly accurate. By contrast, Newton's second approximation is about 11.01883, a result having almost 160,000 times greater error!

While this example was chosen to show off the procedure, it is by no means unusual. Indeed, had one rescaled the constant term in (3) so as to make the root larger, it is easy to show that a proportionately larger x_0 would have produced Newton's Rule approximations having greater relative error whereas the approximations under the new algorithm would have had smaller relative error.

3. The Algorithm. The steps taken in locating a root of (3) are the essence of the new approach but a more efficient way to obtain successive surrogate equations and root estimates is needed to make it an effective algorithm.

THEOREM 1. *If Lemma* 1 *is used to eliminate all of the interior terms of an nth degree polynomial equation* $P(x) = 0$ *using* $R \neq 0$ *as the root estimate, the resulting* (*surrogate*) *equation is*

$$\frac{P'(R)}{nR^{n-1}}x_1^n = \frac{RP'(R)}{n} - P(R). \tag{5}$$

Proof. Straightforward algebra. ∎

Surrogate equations of the type given in (5) are readily solvable by taking nth roots. These n roots are symmetrically located on a circle centered at the origin in the complex plane. One of these roots is the 'new' R needed to iterate the process. The desired root is clearly the one nearest the old R since that choice will move R the least when the new R (here called x_1) is formed. Rather than making comparisons to find it, however, it is preferable to rearrange (5) to obtain:

$$x_1 = R\left[1 - \frac{nP(R)}{RP'(R)}\right]^{\frac{1}{n}}. \tag{6}$$

Formula (6) also involves nth roots but it is better suited to selecting the root to continue the process. Here the desired nth root is the one closest to $(1,0)$ in the complex plane. If these nth roots are called $z_0, z_1, \ldots, z_{n-1}$ in the usual way, it is clear that the only ones that have a chance of being nearest to $(1,0)$ are the first and the last. If the argument of the bracket in (6) is called α, it is easy to see that z_0 is closest to $(1,0)$ if $\alpha < \pi$ and z_{n-1} is closest otherwise.

The new algorithm can now be described. To approximate roots of $P(x) = 0$ (which, without loss of generality, is assumed not to have a root at $x = 0$):

Step 1: If the desired root is known to be near the origin, solve $P(1/z) = 0$ for $z = 1/x$.
Step 2: Determine a preliminary root estimate, $R \neq 0$.
Step 3: Use R and formula (6) to find x_1 and replace R by this number.
Step 4: If R is unsatisfactory as a root estimate, repeat step 3. (Or repeat step 1 if $|R| \ll 1$.)

For reasons to be discussed later, the effectiveness of the algorithm is enhanced if these steps are augmented by a "step 0" involving a translation of roots to eliminate the coefficient of the next-to-the-highest power, and a "step 5" to undo this translation, but these steps are optional and will not be assumed to have been done in the analysis below.

4. Analysis of the Algorithm. Perhaps the first thing to be asked about any root-seeking algorithm is, "What does it do if an approximation is already a root?" A discussion of this question for roots having multiplicity greater than one will be postponed since the answer is complicated both for the new algorithm and for Newton's Rule. It is well known, however, that roots of multiplicity one are fixed points of Newton's iterative process. This is also true of the new algorithm since if R is a root of multiplicity one of $P(x) = 0$, then $P(R) = 0$ and $P'(R) \neq 0$ and hence, by (6), $x_1 = R$.

Another thing to be asked is whether the formulas involved in an algorithm invite 'bogging down' or 'blowing up' or other kinds of erratic behavior. It turns out that both Newton's Rule and formula (6) for the new algorithm have such a defect: both involve division by $P'(R)$, hence neither can tolerate a root estimate at a place where $P'(x) = 0$. Such a root estimate is fatal for Newton's Rule but it is surprisingly easy to correct for the new algorithm! When programmed for a computer, the new algorithm can be modified so that if $P'(R) = 0$, the next value of R can be chosen to be a (random) distant point in the complex plane. As we shall see, such a corrective is ruinous for Newton's Rule but successive iterates of the new algorithm return from the far reaches of the complex plane with such speed that the (revised) algorithm will locate a root after just a few additional steps.

To establish convergence for the new algorithm, first expand (6) to give:

$$x_1 = R - \frac{P(R)}{P'(R)} - \frac{(n-1)}{2R}\left(\frac{P(R)}{P'(R)}\right)^2 + \cdots. \tag{7}$$

The first two terms of (7) are recognizably Newton's formula for x_1, but we must also show that the other terms aid convergence rather than compromise it.

CASE I. *Isolated Roots*. Suppose $x_0 = R$ is the initial estimate for a root of multiplicity one of the equation

$$P(x) = a_0x^n + a_1x^{n-1} + a_2x^{n-2} + \cdots + a_n = 0$$

that is actually at $x = R + Re$. (Note that e is a measure of the relative error made in estimating $(R + Re)$ by R.) Suppose, further, that e is small enough so that $P'(x) \neq 0$ near the root, in particular, so that $P'(R) \neq 0$. In this notation Newton's Rule yields

$$x_1 = R + Re + \frac{P''(R)}{P'(R)}\frac{(Re)^2}{2!} + \cdots, \tag{8}$$

whereas (7) yields

$$x_1 = R + Re + \left[\frac{P''(R)}{P'(R)} - \frac{(n-1)}{R}\right]\frac{(Re)^2}{2!} + \cdots. \tag{9}$$

If we calculate the first and second derivatives of $P(x)$ at $x = R$ and insert them into these formulas, the coefficients of $e^2/2$ in Newton's Rule and the new algorithm are found to be, respectively,

$$\frac{n(n-1)a_0R^n + (n-1)(n-2)a_1R^{n-1} + (n-2)(n-3)a_2R^{n-2} + \cdots}{na_0R^{n-1} + (n-1)a_1R^{n-2} + (n-2)a_2R^{n-3} + \cdots} \tag{10}$$

and

$$-\frac{a_1(n-1)R^n + 2a_2(n-2)R^{n-1} + 3a_3(n-3)R^{n-2} + \cdots}{a_0nR^n + a_1(n-1)R^{n-1} + a_2(n-2)R^{n-2} + \cdots}. \tag{11}$$

The presence in these expressions of quantities that are as yet unspecified limits one to qualitative comparisons but, supposing the dominant terms in (10) and (11) to be those having the highest powers of R, the coefficient of e^2 in Newton's Rule would then be larger than the corresponding coefficient for the new algorithm by a factor of approximately $n|Ra_0/a_1|$, provided that $a_1 \neq 0$. Since a_1/a_0 is the negative of the sum of the roots, this factor simplifies to $|R|/|C_1|$, where C_1 is the centroid of the roots. (The subscript on C indicates that this is the centroid of the first power of the roots. Similar formulas to be developed below will involve centroids of higher powers of the roots.) From this it follows that the new algorithm is likely to require fewer iterative steps than Newton's Rule to locate roots of large modulus to a given degree of accuracy. (Usually far fewer because of the way errors vanish in second-order algorithms.)

It should be pointed out that we are speaking of 'likely' errors arising from 'large' roots of 'typical' equations. Almost certainly equations exist having coefficients that defeat the natural advantages possessed by the new algorithm even for large roots. (Every algorithm has its nemesis!) However, the new algorithm has an additional feature that makes it effective for smaller roots than the foregoing analysis might suggest.

A closer examination of (10) shows that if $|R| \gg 1$ and an error $E = Re$ occurs at a given stage in computations with Newton's Rule, the dominant error term at the next stage is of the order of $ne^2|R|$ because the leading terms in both the numerator and the denominator of (10) are multiples of a_0 which cannot vanish. On the other hand, a_0 does not appear in the numerator of (11). Its leading term involves the first non-vanishing coefficient *after* a_0 and, for each such coefficient that is missing, the new algorithm appears to benefit by another factor of R. Actually, it doesn't benefit quite that much because of offsetting effects from coefficients that do survive, but there is a residual advantage that we will now investigate.

Suppose that coefficients a_1 through a_{k-1} are all zero but that $a_k \neq 0$. In that case the ratio of the second order error terms is approximately $|na_0R^k/(ka_k)|$. This formula is useful to investigate errors in specific equations, but to see how the ratio behaves in general we must digress to examine a somewhat obscure result from the theory of equations that gives the sums of *powers* of roots of polynomial equations.

It is well known that S_1, the sum of first powers of the roots, is given by $(-a_1/a_0)$. Less well known is the fact that for arbitrary $k \leqslant n$ the sum of kth powers S_k, is always a rational function of $a_0, a_1, \ldots, a_k$. The first few formulas of this type are:

$$S_1 = -A_1,$$

$$S_2 = A_1^2 - 2A_2,$$

$$S_3 = -A_1^3 + A_1A_2 - 3A_3,$$

where $A_k = a_k/a_0$. Further formulas like these may be found in [2]. The general formulas are complicated but in our situation a_1 through a_{k-1} all vanish, and in that case S_k simplifies to $(-kA_k)$, that is, to $(-ka_k/a_0)$.

We return to the main argument and suppose that a_k is the first non-zero coefficient following a_0. (By incorporating the translation of roots mentioned as an optional "step 0" when describing the algorithm, one can insure that k is at least two.) Since $S_k = -ka_k/a_0$ under these assumptions, the expression found for the ratio of the second degree error term in Newton's Rule to the similar term for the new algorithm simplifies to $|R^k/C_k|$, where C_k is the centroid of the kth powers of the roots. This estimate shows that, for polynomial equations having one or more "missing" coefficients immediately following the term of highest degree, the new algorithm should be much better than Newton's for finding roots having moduli that are large compared to the others, and that it should be particularly impressive whenever a given iterate is far from the actual roots. The estimate also shows that if any of the roots are complex, the ratio of errors is almost certain to be large even for root estimates having small modulus, because then cancellation occurs in S_k before the centroid C_k or its modulus are computed. In many cases the ratio being discussed can be enormous. In fact, it was a combination of effects discussed in this paragraph that led to the overwhelming victory of the new algorithm over Newton's Rule in the example discussed in the Introduction.

Formulas (10) and (11) also suggest that the new algorithm is likely to lose its advantages over Newton's Rule if $|R| \approx 1$ and to be inferior to it if $|R| \ll 1$. These tendencies can be demonstrated but they aren't uniform. (The presence of other roots sometimes seems to 'confuse' one algorithm more than the other.) Even when present they do not constitute a decisive advantage for Newton's Rule since, as was noted, the case $|R| \ll 1$ can always be converted to the case $|R| \gg 1$ by solving $P(1/z) = 0$. (This trick is also useful if the polynomial has more missing terms near a_n than it does near a_0.)

CASE II. *Multiple roots.* If R is equal to a multiple root of the original equation, both $P(R)$ and $P'(R)$ vanish and hence, by Theorem 1, the resulting surrogate equation degenerates to $0 = 0$. Under the same conditions, Newton's rule produces the improper form $0/0$. These forms show that roots of multiplicity $m > 1$ are not (ordinary) fixed points either of Newton's Rule or the new algorithm. However, since precise equality between estimates and the corresponding roots is a rarity during machine computation, it is more useful to see what happens if R is near—but not precisely at—a multiple root.

Consider the new algorithm first. While the surviving coefficients of the surrogate equation tend to be small if R is near a multiple root, Theorem 1 will still locate the root if enough significant figures can be retained during the computation. To see why, suppose that $P(x)$ has a root of multiplicity m at $(R + Re)$. Then

$$P(x) = [x - (R + Re)]^m Q(x) \tag{12}$$

for some $Q(x)$. Initiating the search process at $x = R$ and assuming that R is close enough to the actual root that the expression $Q(R) \neq 0$, one finds from Theorem 1 that

$$x_1 = R + \frac{Re}{m} + \left[\frac{Q'(R)}{Q(R)} - \frac{n-1}{2R}\right]\left(\frac{Re}{m}\right)^2 + \cdots \tag{13}$$

provided e is small enough. This shows that if one starts at $x = R$, the next iteration will move approximately one-mth of the distance towards the actual root. Since this acts to correct the first power of e, the algorithm is functioning as a first-order rule. Newton's Rule in its primitive form also behaves this way. In fact, to all intents and purposes, its convergence is the same as that shown in (13) since higher order terms play almost no role and Newton's x_1 is identical to (13) through first-order terms.

Since first order convergence is maddeningly slow, special methods are often used to locate multiple roots. For example, multiple roots must also be roots of $P'(x) = 0$ and therefore such

roots are also zeros of the H.C.F. of $P(x)$ and its derivative. It follows that multiple roots are theoretically removable but numerical considerations sometimes limit the effectiveness of this approach. For this reason a number of techniques have been devised to help Newton's Rule find multiple roots directly. While variations of some of these techniques might also apply to the new algorithm, this hasn't yet been done so Newton's Rule—suitably modified—is presently better for finding multiple roots. It should be noted, however, that the ordinary version of Newton's Rule can behave as though multiple roots are present even when none are! In particular, it always does this whenever a root estimate is far from the roots. For example, when seeking roots of equation (2) which has eight isolated roots, initiating Newton's Rule at (100, 200) produces $x_1 \approx (87.5, 175)$; that is, it acts almost as though the equation has a root of multiplicity eight located at the origin. In fact, in that example as late as the 17th iteration, Newton's Rule is still operating in a predominantly first-order fashion! On the other hand, when the new algorithm is initiated far from roots—multiple or not!—it's first few approximations approach the roots very rapidly since it moves at first by taking nth roots. Having been brought close to the roots in just a few steps, it is better able to distinguish closely-spaced roots from multiple ones.

The propensity of Newton's rule to 'see' multiple roots when none exist might not seem serious since presumably one could take pains to initiate the search close to anticipated roots as all textbooks advise. This advice isn't always helpful, however, for if one is unlucky at picking a starting point, a later iterate could land near a place where the first derivative vanishes. When that happens, the next iterate will be far from any of the roots and the pseudo-multiple-roots phenomenon will take over to impede return to the vicinity of the roots. What is worse, there doesn't seem to be any way to detect when such a disaster is about to occur! For example, when one initiates the solution of (2) at (100, 100). Newton's Rule appears to be producing a convergent sequence of approximations and then—after 17 iterations—a root approximation near (.474, .474) yields an x_{18} having a modulus of almost 500 million! While the new algorithm is also capable of sudden leaps to the nether reaches of the complex plane, it finds its way back quicker because it employs nth roots in doing so.

5. Experimental Results. A hand calculator is adequate to find roots of equations like (3) using the new algorithm but to test its effectiveness versus Newton's on equations of higher degree having either real or complex coefficients, the author wrote a computer program that starts both algorithms at a common point and then compares their iterates after one or more steps. In this program, care was taken to reduce the risk of spurious comparisons, e.g., sections of the program embodying different algorithms call the same subprograms whenever possible. On the author's IBM-PC, this program seldom takes more than half a second per iteration on equations up to 20th degree. (The constraint on degree is dictated less by time considerations than by problems with underflow in the value of the bracket in (6) when starting points or later iterates have huge moduli.)

Since the new algorithm requires the computation of everything that Newton's algorithm needs and then uses this information in more complicated ways, it is inescapably slower. The disadvantage will depend, in part, on how the program is written. In the Pascal program written by the author, timing involving batches of 100 iterations showed that the new algorithm took approximately three times as long as Newton's Rule per iteration on cubic equations but that this ratio was reduced to about 3 : 2 for 20th degree equations. Since both algorithms require two polynomial evaluations per iteration, the author attributes this convergence to the fact that the time required for these evaluations looms larger as the degree rises.

Of course the time required by an algorithm to do one iteration isn't as important as the time required to reach a pre-determined level of accuracy on typical problems. To investigate this, the author conducted some three hundred root-finding 'races' between the new algorithm and Newton's Rule. More work of this kind needs to be done but the author can report:

* In many cases the algorithms located different roots when started from the same initial point.

* The new algorithm almost always won these races when one or more of the conditions listed in the Introduction to this paper were present. In particular, it *always* found large roots quicker when started at points having moduli one or more orders of magnitude greater than those of any of the actual roots.
* Newton's Rule usually won when the roots were small in modulus (for reasons we have examined) and, to give it its due, it also seemed better able to decide which root to find first when initiated inside regions containing many roots.
* While only Newton's Rule suffered from the quasi-multiple-roots phenomenon, both algorithms were exasperatingly slow in the presence of (true) multiple roots. The new algorithm was slightly faster at displaying the first-order behavior that betrays the presence of multiple roots but only because it tended to get close enough to be affected a little quicker. There seemed to be no qualitative difference in behavior when both algorithms were started at points that were relatively close to multiple roots.
* The new algorithm did seem to be more steerable in the sense that if the process of locating one root of an equation suggested the general location of another, the new algorithm would usually find it when initiated some distance away in the suggested radial direction. Newton's Rule, by contrast, seemed far less predictable, often returning to roots already found from new initiation points.

One further experiment deserves mention. Since the new algorithm creates surrogate equations by replacing the interior terms of $P(x)$ with pairs of terms that are only approximately equivalent, attempts were made to destabilize the new algorithm by using polynomial equations with interior coefficients that were much larger than the end coefficients. None of these experiments completely defeated the new algorithm but some extremely distorted polynomial equations did induce something close to oscillation between points in the complex plane. This behavior never occurred for the new algorithm in experiments with more normally proportioned equations but oscillation (or near oscillation) happened fairly frequently with Newton's Rule.

6. Extensions and Generalizations. As written, Lemma 1 provides for the replacement of terms involving the pth power of x by terms involving the nth and the zeroth powers of x. The choice of image powers is readily generalized as follows:

LEMMA 2. *If* $|x - R| \ll |R|$ *and* $a \neq b$, *then* $\left(\frac{x}{R}\right)^p \approx \frac{(p-b)}{(a-b)}\left(\frac{x}{R}\right)^a + \frac{(a-p)}{(a-b)}\left(\frac{x}{R}\right)^b.$

Proof. This result can be shown by a Taylor's series expansion similar to that used in the proof of Lemma 1. It can also be shown by using Lemma 1 twice—once with $n = a$ and once with $n = b$—followed by an elimination of the constant terms that result. ■

Substitutions of the sort described in Lemmas 1 and 2 have other uses besides solving equations but, to stick to equation-solving, one extension of the ideas of this paper involves noticing that the constants a, b, and p in Lemma 2 are entirely arbitrary. As a result, the techniques of this paper also apply to polynomial-like equations having non-integer exponents.

To illustrate a variation of equation-solving using Lemma 2, consider the following equation:

$$x^6 + 6x^4 - 1859407 = 0. \tag{14}$$

This equation has a root at $x = 11$ that we estimate using $x_0 = 10$. Until now, when solving such equations we would have used Lemma 1 to split the 4th degree term into a 6th degree component and a constant term in order to obtain a surrogate equation that would be easy to solve. Lemma 2 gives us the alternative of splitting this term into components of 6th and 3rd degree instead, thereby creating a surrogate equation that is quadratic in the variable x^3. Since this splitting doesn't move the term quite so far, it is reasonable to hope that the resulting equation might yield a more accurate value of R.

While a better way to introduce quadratic surrogate equations will be discussed below, the

present approach is an improvement. Solving (14) the old way, one obtains $x_1 \approx 10.997$. Using Lemma 2, one obtains the surrogate equation

$$1.02x_1^6 + 40x_1^3 - 1859407 = 0,$$

which yields $x_1 \approx 10.9992$, an approximation having one-fourth as much error!

Lemma 1 can also be modified in another direction. Certain problems (e.g., finding the yield to maturity of investments) give rise to equations of the form

$$(a + bx)^n Q(x) + R(x) = 0,$$

where $Q(x)$ and $R(x)$ are polynomials. While such equations can be reduced to (ordinary) polynomial equations, the following theorem makes that unnecessary:

THEOREM 2. *If* $|x - R| \ll |R|$, *then* $(1 + x)^n \approx \left(1 + \frac{1}{R}\right)^{n-1} x^n + (1 + R)^{n-1}$.

Proof. A Taylor's expansion argument similar to that of Lemma 1. ∎

As an example of the use of Theorem 2, consider the equation

$$(1 + x)^{10} + 4x^{10} - 99528676 = 0. \tag{15}$$

This equation has a root at $x = 5$ that we will estimate by $x_0 = 4$. Using Theorem 2 with $R = 4$, we see that

$$(1 + x)^{10} \approx (1.25)^9 x^{10} + 5^9$$

and hence (15) goes over into the surrogate equation

$$11.4505806x_1^{10} = 97575551. \tag{16}$$

Solving (16) yields $x_1 \approx 4.93$. While this answer isn't very accurate, notice that the process could be iterated using $R = x_1$. Besides, accuracy is relative! Newton's Rule applied to (15) with $x_0 = 4$ yields approximately 6.85, an answer having about 27 times greater error!

To obtain generalizations of a different sort, note that to this point every substitution has involved replacing one expression in x by two others. A natural question to ask is whether one could further improve accuracy replacing one power of x with three others.

There are many ways to derive a three-replacement formula, but it is convenient to obtain this formula and others like it (including Lemmas 1 and 2) as part of one general theory. To do this, consider the following problem: Find A, B, and C such that x^p and $(Ax^i + Bx^j + Cx^k)$ have the same zeroth, first, and second derivatives at $x = 1$. Solving the appropriate simultaneous linear equations, one finds immediately that

$$A = \frac{V(p, j, k)}{V(i, j, k)}, \quad B = \frac{V(i, p, k)}{V(i, j, k)}, \quad \text{and} \quad C = \frac{V(i, j, p)}{V(i, j, k)},$$

where $V(a, b, c)$ designates Vandermonde's determinant $\begin{vmatrix} 1 & 1 & 1 \\ a & b & c \\ a^2 & b^2 & c^2 \end{vmatrix}$. It follows that

$$x^p \approx \frac{V(p, j, k)}{V(i, j, k)} x^i + \frac{V(i, p, k)}{V(i, j, k)} x^j + \frac{V(i, j, p)}{V(i, j, k)} x^k. \tag{17}$$

If one replaces x by (x/R) in (17), the behavior at $x = 1$ goes over into similar behavior at $x = R$ and therefore we have also shown:

LEMMA 3. *If* $|x - R| \ll |R|$ *and* $V(a, b, c)$ *is Vandermonde's determinant, then*

$$\left(\frac{x}{R}\right)^p \approx \frac{V(p, j, k)}{V(i, j, k)}\left(\frac{x}{R}\right)^i + \frac{V(i, p, k)}{V(i, j, k)}\left(\frac{x}{R}\right)^j + \frac{V(i, j, p)}{V(i, j, k)}\left(\frac{x}{R}\right)^k.$$

It is obvious that Lemma 2—and hence, indirectly, Lemma 1—could also have been derived in this manner. The extension of these ideas to more than three replacement terms is also obvious.

One way to obtain surrogate equations from Lemma 3 is to set $k = 0$, $j = m$, and $i = 2m$ in Lemma 3. If the resulting approximation formula for x^p is used on a general polynomial equation $P(x) = 0$, one obtains the quadratic surrogate equation

$$a\left(\frac{x}{R}\right)^{2m} + b\left(\frac{x}{R}\right)^{m} + c = 0,$$

where

$$\begin{aligned} a &= (1-m)RP'(R) + R^2P''(R), \\ b &= -2\left[(1-2m)RP'(R) + R^2P''(R)\right], \\ c &= 2m^2P(R) + (1-3m)RP'(R) + R^2P''(R). \end{aligned} \tag{18}$$

If one uses (18) instead of (6) to determine new values of R, a generalization of the new algorithm is created. The author incorporated this generalization, Newton's Rule, and the algorithm discussed earlier in a computer program designed to study all three. The generalized version was exceedingly slow—about ten times slower than Newton's Rule on polynomial equations of low degree—but it often found root estimates accurate to a dozen or more places after just two or three iterations. Both its derivation and its behavior under testing lead the author to conjecture that it is a third-order rule.

7. Lingering Questions. Probably the first thing that enters most people's heads upon hearing the words "Newton's Rule" is not Newton's iteration formula but rather the geometric figure showing a tangent line at $(x_k, f(x_k))$ crossing the axis to determine x_{k+1}. While the conviction that figure gives that "it has to work" only applies to searches for real roots, the reader may be wondering what the geometry associated with the algorithms discussed in this paper looks like in the real case.

To answer this question, one further corollary is needed. Because of the way the terms that replace x^p match it in function value and one or more derivatives at $x = R$, we have

COROLLARY. *If $P(x) = 0$ is replaced by the surrogate equation $Q(x) = 0$ using Lemma 1 or Lemma 2, then $P(R) = Q(R)$ and $P'(R) = Q'(R)$. If $Q(x)$ was obtained using Lemma 3, then both of these equations hold and furthermore $P''(R) = Q''(R)$.*

Using this result the geometry for both versions of the new algorithm becomes clear. Rather than replacing $P(x)$ with a (surrogate) line that is tangent at $x = R$ as is done in Newton's Rule, the first version of the new algorithm uses a surrogate tangent curve having the same degree as $P(x)$. The second version uses a curve that is quadratic in a power of x. While this quadratic surrogate curve may or may not match $P(x)$ in degree, it does agree with it through second derivatives at $x = R$. That the places these surrogate curves cross the x-axis "ought to be" good root estimates has considerable intuitive appeal.

The reader might also be wondering how it is that mathematical structures as ill-conditioned as polynomial equations can tolerate the abuse given them by the algorithms of this paper and still produce usable root estimates. Actually the success of the method *is* somewhat mysterious but it isn't magical, i.e., ill-conditioning doesn't mean it must fail. To see this, consider the family of surrogate equations that can be generated from a polynomial equation using Lemma 2 with R set equal to one of the roots. This family has the cardinality of the continuum, since Lemma 2 permits one to range upward or downward in degree and to eliminate part or all of any given interior term. But all of the equations in this family have the same value at $x = R$, hence R is a root of every equation in the family. This dramatizes something that can be overlooked when discussing ill-conditioning of polynomial equations, namely, ill-conditioning is a set-theoretic concept rather than a promise of trouble for a particular root. In the case under discussion, the

set of roots *is* violently affected by Lemma 2—even the number of roots varies across the surrogate family!—yet the favored root at $x = R$ isn't affected at all! In fact, in spite of ill-conditioning, all of the curves of the surrogate family even have the same slope at $x = R$.

8. Concluding Comments. While the generalization of the new algorithm that was conjectured to be a third-order algorithm is interesting, it is not proposed as a serious alternative to Newton's Rule for solving polynomial equations, because it is too slow. On the other hand, the first version is competitive with Newton's Rule in most cases and is markedly superior to it under the circumstances given in (1).

The author is not recommending that Newton's Rule be discarded in favor of the new algorithm but rather that it be complemented by it. Numerical analysts will readily appreciate the advantages of running both algorithms in parallel when solving polynomial equations. While a two-algorithm program would take more time to run than one containing either algorithm alone, the extra time would seldom be wasted. Often one algorithm will find one root while the other finds a second and sometimes one algorithm will locate a root while the other 'bogs down' or 'blows up'. In such situations, using two algorithms would save time rather than waste it. Finally, even in cases where both algorithms approached the same root, having a second estimate to help determine accuracy might still save time by permitting an earlier cessation of iteration.

References

1. G. E. Forsythe, Pitfalls in computation, this MONTHLY, 77 (1970) 943–944.
2. Burnside and Panton, Theory of Equations, Vol. 1, Dover Publications, New York, 1960, pp. 165–167.

A STOCHASTIC CHARACTERIZATION OF THE SINE FUNCTION*

BENJAMIN KEDEM
Department of Mathematics, University of Maryland, College Park, MD 20742

1. Summary. It is shown that conditions based on level crossings and related quantities can ensure that a random sequence contains sinusoidal components. The results are first motivated by considerations of data-reduction in signal analysis, and then formulated precisely for stationary Gaussian sequences. An analogy is drawn with a characterization of the sine function in the nonrandom case.

2. Motivation. The problem dealt with in this paper derives from an important problem of modern technology: the signature problem. To understand the problem, consider the real life example of an operating automobile engine monitored by an electronic device which records the engine's vibration as an oscillating signal. This oscillating signal, regarded as a function of time characterizing the engine's operating condition, is called the engine's "signature". The signature problem is generally to use the signature to tell whether the engine is functioning properly or malfunctioning, and to detect abnormalities such as cracks in the engine block and parts. Signature analysis is regularly performed on various sensitive parts of aircraft which may develop cracks or fractures as a result of metal fatigue.

Benjamin Kedem: I received my Ph.D. in Statistics at Carnegie-Mellon University and joined the Mathematics Department at the University of Maryland in 1975. My research interests lie in time series analysis and in particular zero-crossings analysis. My introduction to signatures is due to the late Dr. I. N. Shimi who was with the Air Force Office of Scientific Research, and who kindly suggested to me fitting this topic into my research plans. I would like to dedicate this article to his memory.

*Research supported by Grant AFOSR 82-0187.

Fig. 1 displays several computer simulated time signals that resemble in appearance real signatures such as those of combustion engines or other man-made systems subjected to repeated testing. The graphs display periodic or nearly periodic oscillation which is suggestive of a superposition of purely periodic components. It might naively be hoped by visual inspection of such graphs to identify the periodic components. Since the periodic components of signatures in engineering applications often have physical interpretations and are of great importance, a more automatic and reliable method of analysis is called for. In order to analyze signatures with the help of digital computers, the continuous time signals are sampled, digitized, and recorded at discrete instants. A signal may be sampled every second or every tenth of a second or even every millisecond. In this paper, we assume sampling is done at regular time intervals Δt, and without loss of generality take $\Delta t = 1$. The graphs in Fig. 1 are constructed from sampled values, but the points are connected by straight line segments to give the impression of continuous curves.

The signal is usually not encoded or analyzed in its raw form, but rather transformed and summarized in some convenient form in which ideally important features of the signal have been enhanced. Such reduction is essential for very long data records. One way to achieve data reduction in signature analysis is to replace the sampled series by counts of zero- or other level-crossings. More generally, Kedem and Slud (1982) discussed the possibility and consequences of reducing the original data on a signal to counts of zero-crossings by the signal and its finite-differences. These counts, called higher-order crossings, are defined as follows. Suppose a discrete-time signal $\{Z_t\}$, that is,

$$\ldots, Z_{-1}, Z_0, Z_1, \ldots, Z_N, \ldots$$

has been recorded. The number of crossings of the time-axis counted in the graph of the series

$$Z_1, \ldots, Z_N$$

when the N points (j, Z_j) are connected by straight line segments, is denoted by D_1. Next, the number of axis-crossings by the graph of the differenced signal

$$Z_1 - Z_0, Z_2 - Z_1, \ldots, Z_N - Z_{N-1}$$

when the N points $(j, Z_j - Z_{j-1})$ are joined by line segments, is denoted by D_2. Note that D_2 is, except for end effects, the number of peaks and troughs in the original undifferenced sequence $Z_1, \ldots, Z_N$. D_1 and D_2 are the first two higher order crossings. In general, let ∇ denote the difference operator: $\nabla Z_t \equiv Z_t - Z_{t-1}$ and $\nabla^k Z_t \equiv \nabla(\nabla^{k-1} Z_t)$. Then the number of axis-crossings by the graph of

$$\nabla^{k-1} Z_1, \nabla^{k-1} Z_2, \ldots, \nabla^{k-1} Z_N,$$

when the N points $(j, \nabla^{k-1} Z_j)$ are connected by line segments, is denoted by D_k and is called the (number of) higher order crossings of order k. The vector of higher order crossings $\{D_1, D_2, D_3, \ldots\}$ is taken to represent the oscillatory information in the signal. Kedem and Slud (1982) argue that in order to achieve an effective data reduction by higher order crossings, it is sufficient to consider only very few D_k's, perhaps only the first ten. The main problem of this paper can now be formulated in terms of higher order crossings by taking a closer look at Fig. 1.

Fig. 1 displays the first ten higher order crossings $D_1, \ldots, D_{10}$ for four examples of series with $N = 200$. In (a), where the signal is a pure sinusoid with angular frequency 0.3, the counts are

$$D_1 = D_2 = \cdots = D_{10} = 19.$$

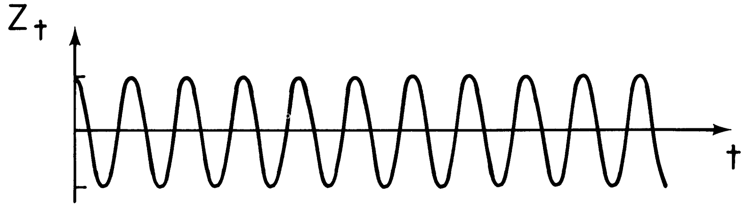
All the D_j are equal, and

$$\pi D_1/(N-1) = 0.2999$$

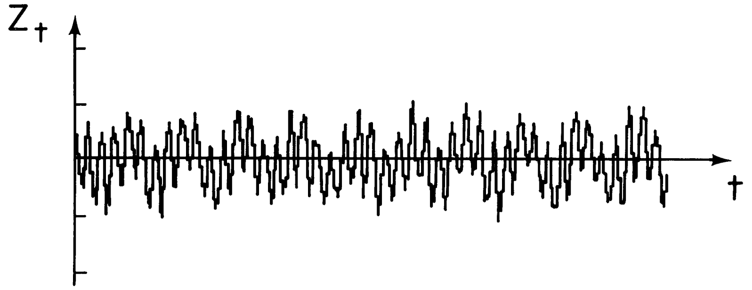
is very close to 0.3, the frequency of the sinusoid.

In case (b),

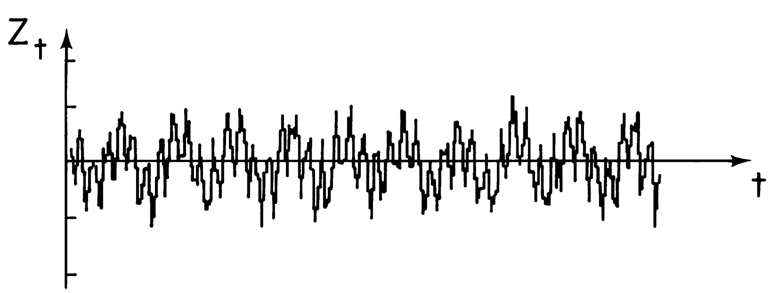
$$D_1 = D_2 = D_3 = 79$$



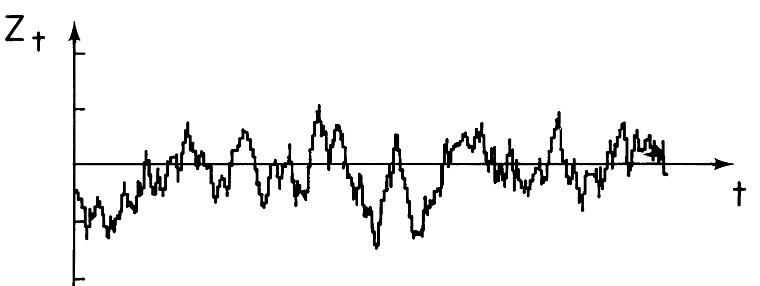
(a) $Z_t = \cos(0.3t)$.
Higher order crossings $\{D_j\}_{j=1}^{10} = \{19, 19, 19, 19, 19, 19, 19, 19, 19, 19\}$.



(b) $Z_t = \cos(0.3t) + 2\cos(1.25t) + 0.2\sin(1.5t) + u_t,\ u_t \sim U(-\frac{1}{4}, \frac{1}{4})$.
Higher order crossings $\{D_j\}_{j=1}^{10} = \{79, 79, 79, 81, 101, 127, 149, 165, 166, 169\}$.



(c) $Z_t = 2.2\cos(0.3t) + 2.8\cos(1.25t) - 1.5\sin(2.6t) + u_t,\ u_t \sim U(-1, 1)$.
Higher order crossings $\{D_j\}_{j=1}^{10} = \{77, 106, 144, 161, 165, 165, 165, 165, 165, 165\}$.



(d) $Z_t = 0.8Z_{t-1} + u_t,\ u_t \sim U(-\frac{1}{2}, \frac{1}{2})$.
Higher order crossings $\{D_j\}_{j=1}^{10} = \{48, 101, 130, 142, 148, 156, 160, 166, 167, 171\}$.

FIG. 1. Computer simulated signatures. $u_t \sim U(a, b)$ means a sequence of computer generated random numbers uniformly distributed in the interval (a, b), for $t = 1, 2, \ldots, 220$.

and

$$\pi D_1/(N-1) = 1.247$$

estimates the frequency 1.25 of one of the sinusoidal components of signal (b) fairly closely.

In (c),

$$D_5 = D_6 = \cdots = D_{10} = 165$$

and

$$\pi D_5/(N-1) = 2.605$$

again estimates closely the frequency 2.6 of one of the signal's sinusoidal components.

In case (d), the D_j increase with j and no two D_j are equal. In this case there are no isolated sinusoidal components.

These examples illustrate that the D_j tend to *increase* as j increases, but that an *equality* among successive D_j values indicates the existence of (isolated) sinusoidal components in a signal. The same phenomenon, observed also in numerous other cases, leads to the general hypothesis that equality among some of the higher order crossings betokens significant sinusoidal components in the signal. In what follows, this experimental finding is made precise under some conditions.

Observe that in each of the cases (b), (c), (d) in Fig. 1, computer generated random numbers u_t have been superposed on purely periodic components for all $t = 1, \ldots, 220$. This superposition of randomness is meant to simulate "noise" generated by an engine itself and by its environment. This is a reasonable way to represent real world signatures, for if the engine were turned off at time T and later re-started, then the new signature of length T will very likely share the general appearance and general characteristics of the first but will not be identical to it. Therefore, in explaining the phenomenon discussed above, the use of tools from the theory of probability seems unavoidable, and the sequence $\{Z_t\}$ will be taken as random.

Our intention is thus to investigate what can be proved about a random sequence such as (b), (c), or (d) in Fig. 1 when some of its higher order crossings (or rather the average values of some of the D_j) coincide. Specifically, the random signal $\{Z_t\}$ will be assumed from now on to be a stationary Gaussian random sequence. The next section defines basic notions in the theory of stationary Gaussian random sequences. Readers familiar with these notions may want to skip ahead to Section 4.

3. Stationary Gaussian random sequences. In order to formulate what we have just seen experimentally in precise mathematical terms, we first define some basic notions from the theory of stationary random sequences defined over a probability space. We shall return repeatedly to the engine example and Fig. 1 for motivation and illustration of abstract concepts. Two good general references on probability and random sequences are the books of Chung (1968) and Karlin and Taylor (1975).

Let S be a set (sample space) of elements s, $\mathscr{A}$ a σ-field of subsets (events) of S, and P a probability measure defined on the elements (sets) of $\mathscr{A}$. The triple $(S, \mathscr{A}, P)$ is called a probability space. A random variable is a real valued function $Z(s)$, $s \in S$, such that for each real z, $\{s\colon Z(s) \leqslant z\} \in \mathscr{A}$. That is, Z is a measurable function on $(S, \mathscr{A})$. The mathematical expectation of Z is defined (Chung (1968), p. 39) as the integral over the sample space with respect to the measure P and is denoted by

$$EZ \equiv \int_S Z(s)\,P(ds).$$

The probability distribution function F of a random variable Z is a nondecreasing function of a real variable defined from the set function P by

$$F(z) = P\{s\colon Z(s) \leqslant z\}.$$

Usually the random element s is dropped in order to simplify the notation, and we write simply

$$F(z) = P(Z \leqslant z).$$

In the engine example, one might think of s as representing the "state" of the engine and its environment and S as the set of all possible states. For fixed t_0, the random variable $Z_{t_0}(s)$ might correspond to a scalar physical measurement made at time t_0 on the engine in state s. For example, s could be "running" or "off" and $Z_{t_0}(\text{running}) = 1$, $Z_{t_0}(\text{off}) = 0$. Generally, s embodies the complex mechanism of the engine which causes its signature to reach level $Z_{t_0}(s)$ at time t_0. In this case it is of interest to compute the probability that $Z_{t_0}(s)$ stays below a certain threshold. It is useful to think of the expectation EZ_{t_0} as the average value of $Z_{t_0}(s)$ in the following sense. We start up the engine, finding it in state s_1, and measure $Z_{t_0}(s_1)$ after t_0 time units. Then the engine is turned off, re-started (in state s_2) and measured t_0 time units later to give $Z_{t_0}(s_2)$. After a large number m of such re-start cycles, the resulting numerical average of the m values $Z_{t_0}(s_j)$ approximates EZ_{t_0}.

A random sequence is a countable collection or family of random variables $\{Z_t\}$, $t = 0, \pm 1, \ldots$, defined over the same probability space. A record of actual values observed at $t = 0, \pm 1, \ldots$,

$$\ldots, Z_{-1}(s), Z_0(s), Z_1(s), Z_2(s), \ldots$$

(graphed and connected by line segments for easy visualization) is called a realization. Fig. 1 contains finite realizations of four different simulated random sequences. Or again, each time an engine is started up it may produce a realization of measurements $Z_t(s)$ which we earlier called its signature.

So far we have not assumed any special structure for $\{Z_t\}$, but we do so now by considering the realizations in Fig. 1 as motivation. Each realization there fluctuates about a fixed level, but the general appearance of each realization is repetitious and does not change with time shifts. This observation leads to the following reasonable assumptions: first, that Z_t should behave no differently than $Z_{t'}$, whenever $t \neq t'$ are fixed, at least as far as averages are concerned; secondly, that the relationship between Z_t and $Z_{t'}$ resembles that of the shifted pair Z_{t+k} and $Z_{t'+k}$. Formally, we assume

$$\text{(a)} \quad EZ_t = \text{constant}, \quad \text{for all } t.$$

$$\text{(b)} \quad EZ_tZ_{t+k} = \rho_k, \quad k = 0, \pm 1, \ldots, \quad \text{for all } t.$$

Without loss of generality we always take the constant in (a) as 0, and $\rho_0 = 1$ in (b). ρ_k is a function of the lag k, called the correlation function of $\{Z_t\}$, and is seen to be an even function of k. A random sequence for which (a) and (b) hold is called *stationary*. Thus, the first condition that we impose on $\{Z_t\}$ is that it is a stationary random sequence. Experience shows that this assumption is useful in treating sequences whose realizations resemble those given in Fig. 1.

The correlation function plays an important role in the theory of stationary sequences. Observe that for any constants $\alpha_1, \ldots, \alpha_m$.

$$0 \leqslant E\left|\sum_{j=1}^{m} \alpha_j Z_{t_j}\right|^2 = \sum_{j,k} \alpha_j \alpha_k EZ_{t_j}Z_{t_k} = \sum_{j,k} \alpha_j \alpha_k \rho_{(t_j - t_k)},$$

so that ρ_k is positive semidefinite by definition.

A formal example of a stationary sequence is furnished as follows. Let A, B be orthogonal random variables. That is, $EA = EB = 0$, $EAB = 0$. Also, let $EA^2 = EB^2 = 1$. Then the sequence

$$Z_t = A\cos\omega t + B\sin\omega t, \quad t = 0, \pm 1, \ldots \tag{1}$$

is stationary. This follows since by linearity of the integral $EZ_t = 0$ for all t, and by the orthogonality of A and B

$$EZ_tZ_{t+k} = \cos\omega t \cdot \cos\omega(t+k) + \sin\omega t \cdot \sin\omega(t+k) = \cos\omega k = \rho_k,$$

which is a function of k only. Note that $|\rho_k| \leqslant 1$, which for $\rho_0 = 1$ is a general property of ρ_k which follows from the Schwarz inequality.

So far the only restriction we imposed on the random variables Z_t are some moment conditions, while no assumption was made about probability distributions. A fruitful assumption which will lead us to some insights is the so-called Gaussian assumption, endowing the random variables Z_t with the (joint) normal distribution. The Gaussian assumption means that every linear combination

$$Y = \alpha_1 Z_{t_1} + \cdots + \alpha_n Z_{t_n}$$

for arbitrary time-indices $t_1, \ldots, t_n$, has a normal distribution. Since we assume that $EZ_t = 0$, this means that

$$P(Y \leqslant y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}\,\sigma} \exp(-x^2/2\sigma^2)\,dx,$$

where

$$\sigma^2 = \sum_{i,j=1}^{n} \alpha_i \alpha_j \rho_{(t_i - t_j)}.$$

See Karlin and Taylor (1975, §9.8). In this paper it is assumed throughout that $\{Z_t\}$ is a stationary Gaussian random sequence.

4. The sinusoidal limit. We now have a framework within which we can explain the experimental results in Section 1. The key factor in our theoretical development is the interesting connection between ρ_1 and ED_1 under the Gaussian assumption. Observe that D_1, as a function of $Z_1, \ldots, Z_N$, is itself a random variable which takes on values $0, 1, 2, \ldots, N-1$. Therefore

$$0 \leqslant ED_1 \leqslant N - 1.$$

When ED_1 is close to 0 we expect smooth realizations, and when ED_1 is close to $N-1$, we expect oscillatory realizations. This should have a bearing on ρ_1 which measures the correlation between neighboring Z_t's. The connection is stated in

LEMMA 1. *Let $\{Z_t\}$ be a zero mean stationary Gaussian random sequence with correlation function ρ_k. Then*

$$\rho_1 = \cos\left(\frac{\pi E(D_1)}{N-1}\right).$$

Proof. Let d_t be the indicator of the event

$$\{Z_{t-1} < 0, Z_t \geqslant 0\} \cup \{Z_{t-1} \geqslant 0, Z_t < 0\}.$$

Then $d_t = 1$ if the event occurs and $d_t = 0$ otherwise, and

$$D_1 = d_2 + \cdots + d_N.$$

Consider the pair Z_t, Z_{t+1}. By the Gaussian assumption this pair has a bivariate normal distribution (Karlin and Taylor (1975), p. 14). Therefore

$$P(Z_{t-1} \geqslant 0, Z_t \geqslant 0) = \int_0^\infty \int_0^\infty h(x,y)\,dx\,dy,$$

where

$$h(x,y) = \frac{1}{2\pi\sqrt{1-\rho_1^2}} \exp\left\{-\frac{1}{2(1-\rho_1^2)}\left(x^2 - 2\rho_1 xy + y^2\right)\right\}.$$

Switching to polar coordinates and integrating first with respect to the radial variable leaves

$$P(Z_{t-1} \geqslant 0, Z_t \geqslant 0) = \frac{\sqrt{1-\rho_1^2}}{2\pi} \int_0^{\pi/2} \frac{1}{1-\rho_1 \sin 2\theta}\, d\theta$$
$$= \frac{1}{4} + \frac{1}{2\pi} \sin^{-1}\rho_1.$$

Now by the symmetry of the normal distribution

$$\frac{1}{2} = P(Z_{t-1} \geqslant 0) = P(Z_{t-1} \geqslant 0, Z_t \geqslant 0) + P(Z_{t-1} \geqslant 0, Z_t < 0),$$

and

$$P(Z_{t-1} \geqslant 0, Z_t < 0) = \frac{1}{2} - \left[\frac{1}{4} + \frac{1}{2\pi}\sin^{-1}\rho_1\right]$$
$$= P(Z_{t-1} < 0, Z_t \geqslant 0).$$

Since $Ed_t = P(d_t = 1)$ we have $Ed_t = \frac{1}{2} - \frac{1}{\pi}\sin^{-1}\rho_1$, for all t. But $ED_1 = (N-1)(Ed_t)$ and the result follows by solving for ρ_1. □

With the help of this lemma we finally have what we wanted all along, a connection between oscillatory structure and equality between successive ED_j.

THEOREM 1. *Let $\{Z_t\}$ be a stationary Gaussian random sequence with mean 0 and correlation function ρ_k. Assume $ED_1 > 0$. Then*

(a) *The sequence $\{ED_j\}$ is monotone nondecreasing and bounded,*

$$0 \leqslant ED_1 \leqslant ED_2 \leqslant ED_3 \leqslant \cdots \leqslant (N-1).$$

(b) *If $ED_1 = ED_2$, then $\{Z_t\}$ is a pure sinusoid as in* (1) *with*

$$\omega = \frac{\pi E(D_1)}{N-1}.$$

Proof. Since $\{Z_t\}$ is Gaussian, so is $\{\nabla Z_t\}$ by definition. The number of axis-crossings of $\nabla Z_1, \ldots, \nabla Z_N$, denoted by D_2, will not change if ∇Z_t is multiplied by the constant $1/\sqrt{2(1-\rho_1)}$. But $\{\nabla Z_t/\sqrt{2(1-\rho_1)}\}$ is a stationary Gaussian sequence with mean 0 and correlation function, say, $\rho_\nabla(k)$ where $\rho_\nabla(0) = 1$. By Lemma 1,

$$\rho_\nabla(1) = \cos\left(\frac{\pi E(D_2)}{N-1}\right).$$

From Kedem (1984a), there exists a positive constant K which depends on ρ_1 only such that

$$0 \leqslant E(Z_t - 2\rho_1 Z_{t-1} + Z_{t-2})^2 \leqslant K(\rho_1 - \rho_\nabla(1)). \tag{2}$$

Therefore

$$\cos\left(\frac{\pi E(D_1)}{N-1}\right) \geqslant \cos\left(\frac{\pi E(D_2)}{N-1}\right).$$

Since $0 \leqslant ED_j/(N-1) \leqslant 1$, for all j, and since $\cos(x)$ is monotone decreasing in $[0, \pi]$, it follows that

$$ED_1 \leqslant ED_2.$$

But the relation between D_1 and D_2 is the same as between D_2 and D_3, and between D_3 and D_4, etc. Therefore $ED_2 \leqslant ED_3$, $ED_3 \leqslant ED_4$, and so on, and (a) is proved. However, if $ED_1 = ED_2$,

then (2) vanishes, so that the integrand $Z_t - 2\rho_1 Z_{t-1} + Z_{t-2}$, must itself vanish. Therefore for each t we obtain in this case the equation

$$Z_t - 2\rho_1 Z_{t-1} + Z_{t-2} = 0$$

which is a difference equation whose solution is (1) with period $2(N-1)/ED_1$. □

We see that when ED_1 is equal to ED_2 then the sequence $\{Z_t\}$ must be sinusoidal. Similarly, if $ED_j = ED_{j+1}$, then $\{Z_t\}$ must contain a superposed sinusoidal component with frequency $\pi ED_j/(N-1)$. *Thus the explanation for our experimental results is that in the Gaussian case, at least, the D_j tend to increase on the average, but if they "touch" there is an indication that a purely sinusoidal component is contained in the data.*

In Fig. 1(a), all the D_j of a sinusoid are the same. But what we have proved is exactly the converse! We summarize by saying that under the Gaussian assumption, $\{Z_t\}$ is a sinusoid if and only if $ED_1 = ED_2$.

It should be remarked that Theorem 1 is a special case of a more general result due to Slutsky (1927). The connection with Slutsky's work is discussed in Kedem (1984a).

To conclude the Section, let us have a second look at the sequence $\{ED_j\}$. From Theorem 1, this sequence is monotone and bounded and thus convergent as $j \to \infty$. We shall not attempt a detailed discussion of the limit, but note that if $\{Z_t\}$ is made of superposed sinusoids, then

$$\frac{\pi E(D_j)}{N-1} \to \omega^*, \quad j \to \infty,$$

where ω^* is the highest frequency present among these sinusoids. This convergence means that as j increases, $\nabla^j Z_t$ approaches a pure sinusoid with frequency ω^*.

5. Two examples.

5.1.1. *The sunspot series.* In real applications ED_j is replaced by D_j. Here is an example of an

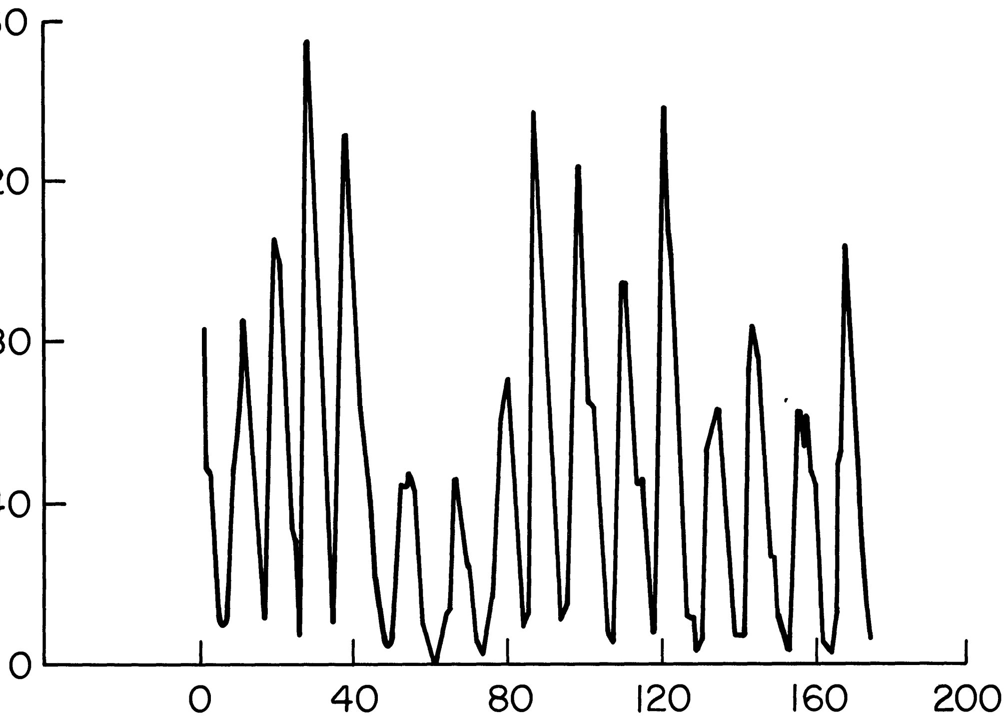


FIG. 2. Graph of the Wolfer's Sunspot Numbers from 1749 to 1924.

application of Theorem 1 to real data. Of course the Gaussian assumption may not hold exactly in practice, but Fig. 1 emboldens us to hope that this assumption is not absolutely needed for a reasonable use of Theorem 1.

Fig. 2 shows the graph of annual sunspot numbers from 1749 to 1924. It is well known that these data contain several periodic components plus possible "noise". The most significant period is known to be a little over 11 years. See Anderson (1971, p. 244 and pp. 659–662) for the actual records and their analysis by conventional techniques. We have applied to these data a certain linear operation (mainly repeated summation; see Kedem (1984b)) which smooths the data and reduces "noise". With $N = 148$, the first eight D_j are $7, 24, 26, 26, 28, 38, 45, 60$, and we see that $D_3 = D_4 = 26$. The frequency which corresponds to $D_3 = 26$ is $\pi \cdot 26/147 = 0.556$, and the corresponding period is $2\pi/0.556 = 11.301$.

5.1.2. *A second order autoregressive scheme*. Consider graph (d) in Fig. 1. The graph looks "periodic" and yet it does not contain isolated periodic components. Early this century it was realized that fitting to such data superposed sinusoids plus "noise" in order to detect periodicities was not a very fruitful practice as this often led to the "discovery" of spurious periods. In 1927, G. U. Yule published a celebrated paper in which he suggested a way to model data such as in graph (d). He considered a damped sine wave in which the damping factor was controlled by a stream of random shocks. The result was a model of disturbed harmonics of the form

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \varepsilon_t, \quad t = 0, \pm 1, \dots,$$

referred to as an autoregressive scheme. This type of model has found many applications in time series analysis and signal analysis. The noise term (random shocks) $\{\varepsilon_t\}$ consists of orthogonal normal random variables with mean 0:

$$E\varepsilon_t\varepsilon_{t'} = \begin{cases} \sigma_\varepsilon^2, & t = t', \\ 0, & t \neq t', \end{cases} \qquad EZ_t\varepsilon_{t'} = 0, \quad t < t'.$$

If the roots of $1 - \phi_1 x - \phi_2 x^2 = 0$ lie outside the unit circle, then $\{Z_t\}$ constitutes a stationary zero-mean Gaussian random sequence (Karlin and Taylor 1975, Chap. 9). This sequence satisfies the hypothesis of Theorem 1.

The orthogonality of $\{\varepsilon_t\}$ implies

$$\rho_1 = \phi_1 + \phi_2 \rho_1,$$
$$\rho_2 = \phi_1 \rho_1 + \phi_2.$$

Also, it is not difficult to see that

$$\rho_2 = -1 + 2\rho_1 - 2\rho_\nabla(1)(1 - \rho_1).$$

Therefore, solving for ϕ_1 and ϕ_2, we find

$$\phi_1 = \frac{2\rho_1(1 + \rho_\nabla(1))}{1 + \rho_1}, \qquad \phi_2 = \frac{-1 + \rho_1 - 2\rho_\nabla(1)}{1 + \rho_1}.$$

Also, since we assume that $\rho_0 = 1$, we have

$$\sigma_\varepsilon^2 = 1 - \rho_1\phi_1 - \rho_2\phi_2,$$

which together with the preceding expressions for ϕ_1, ϕ_2, ρ_2, yields the inequality

$$0 \leqslant \sigma_\varepsilon^2 \leqslant 8(\rho_1 - \rho_\nabla(1)).$$

Hence, when $0 \leqslant ED_2 - ED_1 \to 0$,

$$\phi_1 \to 2\rho_1, \quad \phi_2 \to -1, \quad \sigma_\varepsilon^2 \to 0,$$

and Z_t tends to satisfy a difference equation whose solution is (1). Note that $\sigma_\varepsilon^2 = 0$ implies that ε_t vanishes for every t. It is seen that the equality $ED_1 = ED_2$ means that Z_t becomes an

undamped and undisturbed sine wave in agreement with Theorem 1. The other extreme is the case when

$$ED_1/(N-1) = 1/2 \quad \text{and} \quad ED_2/(N-1) = 2/3.$$

Then $\phi_1 = \phi_2 = 0$ and $Z_t \equiv \varepsilon_t$; in that case, Z_t is all noise. (See also Wold (1965), pp. 36–37.)

6. An analogy with the deterministic case. Roe (1980) has proved the following result which characterizes the sine function.

THEOREM 2. *Let* $\{f^{(n)}\}$, $n = 0, \pm 1, \ldots$, *be a sequence of real valued functions of a real variable* t, $-\infty < t < \infty$, *where*

$$f^{(n+1)}(t) = \frac{d}{dt} f^{(n)}(t).$$

If there exists a real M *such that*

$$|f^{(n)}(t)| \leqslant M, \quad \textit{for all } n, t,$$

then $f^{(0)}(t) = a \sin(t + \phi)$ *for some real constants* a, ϕ.

Let us draw an analogy between the deterministic and random cases as manifested by Theorem 2 and Theorem 1, respectively. In the deterministic case $f^{(0)} \equiv f$, as well as its derivatives and integrals, must be bounded. But in the random case, the sequence $\{Z_t\}$ is Gaussian and therefore not bounded in the strict sense, as the normal distribution is defined over the whole real line. However, the requirement that $EZ_t^2 = \rho_0 = 1$ may be viewed as a condition analogous to the boundedness of f. Next, Roe's proof is based on a demonstration that the Fourier transform of f vanishes everywhere except at $+1$ and -1. However, we have not mentioned any Fourier type analysis in our proof. Still there is an analogy.

Recall that earlier it was shown that ρ_k is positive semidefinite. This implies (Karlin and Taylor (1975), p. 504) that there exists a symmetric probability distribution function F such that ρ_k admits the Fourier representation

$$\rho_k = \int_{-\pi}^{\pi} \cos(k\omega)\, dF(\omega),$$

where the integral is a Stieltjes integral. What can be said of F when $ED_1 = ED_2$? Observe that by the Gaussian assumption

$$\rho_1 = \cos\left(\frac{\pi ED_1}{N-1}\right) = \int_{-\pi}^{\pi} \cos(\omega)\, dF(\omega)$$

and

$$\rho_\nabla(1) = \cos\left(\frac{\pi ED_2}{N-1}\right) = \frac{\int_{-\pi}^{\pi} \cos(\omega)(1-\cos(\omega))\, dF(\omega)}{\int_{-\pi}^{\pi} (1-\cos(\omega))\, dF(\omega)}.$$

Since F is a probability distribution function, there exists a random variable Y with distribution function F. Then

$$\rho_1 = E\cos(Y), \quad \rho_\nabla(1) = \frac{E\cos(Y)(1-\cos(Y))}{1-E\cos(Y)},$$

so that, when $ED_1 = ED_2$, $\rho_1 = \rho_\nabla(1)$ and

$$E\cos^2(Y) = E^2\cos(Y).$$

Hence $E(\cos(Y) - E\cos(Y))^2 = 0$ or

$$\cos(Y) = \cos\left(\frac{\pi E(D_1)}{N-1}\right).$$

Thus Y is a random variable which admits only two values in $[-\pi, \pi]$, say $+\lambda_0$ and $-\lambda_0$. Since F is a symmetric probability distribution function, it must have two jumps, at $+\lambda_0$ and $-\lambda_0$, each of size $1/2$. In other words

$$\begin{aligned} F(\lambda +) - F(\lambda -) &= \frac{1}{2}, \quad \lambda = \pm\lambda_0, \\ &= 0, \quad \text{otherwise.} \end{aligned}$$

This degeneracy of F is analogous to the singularity in the pure-sinusoid characterization of Roe in the deterministic case.

Acknowledgement. The efficient editing help of Eric Slud is gratefully acknowledged.

References

1. T. W. Anderson, The Statistical Analysis of Time Series, Wiley, 1971.
2. K. L. Chung, A Course in Probability Theory, Harcourt, Brace & World, 1968.
3. S. Karlin and H. M. Taylor, A First Course in Stochastic Processes, Academic Press, 1975.
4. B. Kedem, On the sinusoidal limit of stationary time series, Ann. Statist., 12 (1984a) 665–674.
5. ______, Detection of hidden periodicities by means of higher order crossings, Technical Report TR 84–56, Department of Mathematics, University of Maryland (1984b).
6. B. Kedem and E. Slud, Time series discrimination by higher order crossings, Ann Statist., 10 (1982) 786–794.
7. J. Roe, A characterization of the sine function, Math. Proc. Cambridge Philos. Soc., 87 (1980) 69–73.
8. E. E. Slutsky, The summation of random causes as the source of cyclic processes (in Russian), Problems of Economic Conditions, 3 (1927), English translation, Econometrica, 5 (1937) 105–146.
9. H. Wold, Bibliography on Time Series and Stochastic Processes, M.I.T. Press, 1965.
10. G. U. Yule, On a method for investigating periodicities in disturbed series with special reference to Wolfer's Sunspot Numbers, Philos. Trans. Roy. Soc. London Ser. A., 226 (1927) 267–298.

COMMUTATORS AND THEIR PRODUCTS

I. D. MACDONALD
Department of Mathematics, University of Hawaii, Honolulu, HI 96822

1. The Problem. The product of two commutators in an arbitrary group need not be a commutator. The first to establish this fact seems to have been W. B. Fite [2], as long ago as 1902. Since then numerous examples have been published, mostly with interesting special features; see [7], [3], [5], [1], etc.

Nevertheless, the beginner still has a problem. The groups with which he is familiar tend to have small orders or straightforward structures. Products of commutators in such groups tend to be commutators. An impression therefore comes about that "non-commutators" are unusual; maybe even pathological.

It is our aim to persuade this beginner that "non-commutators" exist in an abundance of groups. Moreover, it is easy to construct such groups. The attitude that there are only a few exotic examples, kept in glass cases in some museum of groups, is wrong. We are knee-deep in examples, or will be by the end of this note.

I. D. Macdonald tries to write research papers and even books on mathematics. He also has secret vices; he writes awful plays and has done 100 parachute jumps, the first when he was 45.

Most of our material is quite well known. What we stress is its elementary nature.

By the way, we should have said that a commutator in an arbitrary group G is an element of the form $a^{-1}b^{-1}ab$, usually written $[a, b]$, for some elements a, b of G. The least subgroup $\delta(G)$ containing all the commutators is said to be the commutator subgroup, or derived group, of G. Thus the situation we are interested in is that not every element of $\delta(G)$ is a commutator.

2. Preliminaries. Where would be a good place to seek the groups we want? A good point is the following almost trivial lemma. It is very well known; formal references are [7], [3].

LEMMA 1. *If G is any group and if*

$$|G\colon \zeta(G)|^2 < |\delta(G)|, \tag{1}$$

then there are elements in $\delta(G)$ that are not commutators.

(By $\zeta(G)$ we mean the center of G, the subgroup formed by those elements which commute with every element of G. The symbol $G\colon\zeta(G)$ denotes the quotient group and the vertical bars indicate its order.)

Proof. It is easy to check that if $a, b \in G$ and $y, z \in \zeta(G)$ then

$$[ay, bz] = [a, b]. \tag{2}$$

So all the commutators $[a, b]$ in G are obtained as a, b run through a set of coset representatives of $\zeta(G)$ in G. In effect there are $|G\colon\zeta(G)|$ choices for a, and the same number for b; there are at the very most $|G\colon \zeta(G)|^2$ commutators. The result follows.

The lemma indicates that we should perhaps look at groups G in which $\zeta(G)$ is relatively "very large" and $\delta(G)$ is reasonably "large". Could we perhaps make something out of the case when $\delta(G) \leqslant \zeta(G)$? It would be pleasant if we could because calculation is particularly easy in this situation. That is a consequence of the next result:

LEMMA 2. *In a group G in which $\delta(G) \leqslant \zeta(G)$*

$$[xy, z] = [x, z][y, z], \tag{3}$$

$$[x, yz] = [x, y][x, z], \tag{4}$$

$$(xy)^n = x^n y^n [x, y]^{-n(n-1)/2} \tag{5}$$

for all x, y, z in G and all positive integers n.

Proof. As a token we prove (3):

$$\begin{aligned}[xy, z] &= (xy)^{-1}z^{-1}(xy)z = y^{-1}(x^{-1}z^{-1}xz)z^{-1}yz \\ &= (x^{-1}z^{-1}xz)(y^{-1}z^{-1}yz) = [x, z][y, z].\end{aligned}$$

(4) is similar and (5) requires induction on n.

Thus the commutator function $[x, y]$ is bilinear if $\delta(G) \leqslant \zeta(G)$, by (3) and (4). The significance of (5) is that, if p is prime, the elements of p-power order in G form a subgroup (as the reader should check). This and the two definitions to come are perfectly well-known facts.

DEFINITION. The group G is said to be of class 2 if and only if $\delta(G) \leqslant \zeta(G)$.

DEFINITION. The group G is said to be a p-group (where p is prime) if all its elements have p-power order.

So we reach this conclusion: to find groups of the sort we are after we could do worse than look at finite p-groups of class 2.

3. A matrix construction. If G is generated by elements $a_1, \dots, a_n$ (we write $G = \langle a_1, \dots, a_n \rangle$), then each element a in G can be written as

$$a = a_1^{k_1} \cdots a_n^{k_n} z, \tag{6}$$

where the k's are integers and $z \in \delta(G)$. If G has class 2, then we have in addition, $z \in \zeta(G)$. How about commutators in G of class 2? If

$$b = a_1^{l_1} \cdots a_n^{l_n} y \tag{7}$$

with $y \in \zeta(G)$, then the bilinearity of the commutator function (expressed by (3) and (4)) and a little calculation give

$$[a, b] = \prod_{i<j} [a_i, a_j]^{k_i l_j - k_j l_i}. \tag{8}$$

Thus we see that if G (of class 2) is generated by n elements $a_1, \dots, a_n$, then $\delta(G)$ is generated by the $n(n-1)/2$ commutators $[a_i, a_j]$. This set of generators may not, of course, be minimal. In general it will not be.

But suppose we could construct an n-generator group G of class 2 with $\delta(G)$ requiring no fewer than $n(n-1)/2$ generators. It looks as though Lemma 1 would be applicable, at least if we could make G finite, and we might have a group with "non-commutators".

This motivates our next move, a matrix construction of such a group in which $\delta(G)$ cannot be generated by fewer than $n(n-1)/2$ elements.

Let m be some large positive integer, let I denote the unit $m \times m$ matrix, and define $E(i, j)$ for $1 \leqslant i < j \leqslant m$ to be the matrix with the integer 1 in row i column j, and with 0 as every other entry. If

$$A = I + E(i, j), \quad B = I + E(k, l), \tag{9}$$

then

$$AB = I + E(i, j) + E(k, l) + E(i, j) E(k, l),$$

with $E(i, j) E(k, l) = 0$ unless $j = k$, in which case the product is $E(i, l)$. Further,

$$A^{-1} = I - E(i, j).$$

The reader will easily verify these equations, and similar rules for forming products and inverses of matrices of the form

$$I + q_1 E(i_1, j_1) + \cdots + q_r E(i_r, j_r), \tag{10}$$

where $q_1, \dots, q_r$ are integers. The inverse of (10) is

$$I - q_1 E(i_1, j_1) - \cdots - q_r E(i_r, j_r).$$

Next, commutators. It is clear that $[A, B] = I$ in (9) if $j \neq k$. If

$$A = I + E(i, j), \quad B = I + E(j, k), \tag{11}$$

then we have

$$[A, B] = I + E(i, k). \tag{12}$$

This equation (12), which is basic, can be extended in an obvious way to give the commutator of two elements of the form (10).

By way of an exercise let us now construct a group $G = \langle A_1, A_2, A_3 \rangle$ of class 2 in which $\delta(G)$ requires $3(3-1)/2$ generators. We put

$$\begin{aligned} A_1 &= I + E(1,2), \\ A_2 &= I + E(3,4) + E(2,7), \\ A_3 &= I + E(5,6) + E(2,8) + E(4,9) \end{aligned}$$

with $m = 9$, say. Then

$$[A_1, A_2] = I + E(1,7),$$
$$[A_1, A_3] = I + E(1,8),$$
$$[A_2, A_3] = I + E(3,9),$$

three matrices which clearly generate a central subgroup of G which clearly has no pair of generators.

We leave it to the reader to generalize the above construction to $\langle A_1, A_2, \dots, A_n \rangle$. As a hint, we mention that one possibility (by no means the only one) is to take

$$A_i = I + E(2i - 1, 2i) + \text{other } E\text{'s}.$$

The reader who is still reading will find no enormous difficulty in this. Some skill, however, is needed in arranging the details and writing them down clearly.

4. Examples. This is the point at which one experiences a feeling of regret for not having taken all matrix entries to be in $\mathbb{Z}_p$, the field of integers modulo p, for all the arguments and construction still work, giving finite groups, which are what we want.

So let's now consider groups G of matrices in which the matrix entries are from $\mathbb{Z}_p$. Then matrices of the form (9) have order p. Matrices of the form (10) may not have order p (if $p > 2$) but (by (6)) their pth powers lie in $\delta(G)$, which is central; we made $\delta(G)$ central by the way we chose the A_i.

We sum up.

PROPOSITION 1. *To each natural number $n > 1$ and each prime p there exists a finite p-group $G = G(n, p)$ of class 2 with the property that $G/\zeta(G)$ and $\delta(G)$ are elementary abelian p-groups of order p^n and $p^{n(n-1)/2}$ respectively.*

Two remarks. First, we have $\delta(G) = \zeta(G)$, a fact which we have not proved (though here the reader may care to test his skill) because we do not need it. Secondly, the matrix group that we constructed is one particular G which satisfies the proposition, but there is no reason to think it is the only one; in fact it is far from being unique.

Now recall Lemma 1. We have elements which are not commutators in the commutator subgroup of our $G(n, p)$ if $2n < n(n-1)/2$, that is if $n \geqslant 6$.

PROPOSITION 2. *If $n \geqslant 6$, then there are products of commutators in $G(n, p)$ that are not themselves commutators.*

Again, to keep things simple, we have not gone for the best possible result. In fact, Proposition 2 is still true if "$n \geqslant 6$" is replaced by "$n \geqslant 4$." Indeed, $G(4, p)$ even has a factor group of order p^8 in which non-commutators figure. The order 2^8 is the least possible and is the order of W. B. Fite's group. See also [**8**], page 78, problem 5.

One very last remark. It would not be at all hard to extend Lemma 1 and Proposition 2 as follows. Given any N, we can find an n such that there are elements in $\delta G(n, p)$ which are not the product of any N commutators.

Acknowledgment. I am pleased to thank Dr. A. D. Sands of the University of Dundee for encouraging me to assemble elementary material about commutators, and I am grateful to various "general" audiences who have patiently listened to versions of this Note.

References

1. Phyllis Joan Cassidy, Products of commutators are not always commutators: an example, this MONTHLY, 86 (1979) 772.

2. W. B. Fite, On metabelian groups, Trans. Amer. Math. Soc., 3 (1902) 331–353.

3. P. X. Gallagher, The generation of the lower central series, Canad. J. Math., 17 (1965) 405–410.

4. Robert M. Guralnick, On a result of Schur, J. Algebra, 59 (1979) 302–310.
5. I. M. Isaacs, Commutators and the commutator subgroup, this MONTHLY, 84 (1977) 720–722.
6. Hans Liebeck, A test for commutators, Glasgow Math. J., 17 (1976) 31–36.
7. I. D. Macdonald, On a set of normal subgroups, Proc. Glasgow Math. Assoc., 5 (1962) 137–146.
8. Ian D. Macdonald, The Theory of Groups, Clarendon, Oxford, 1968.

A VARIATION ON THE LANDAU PROBLEM

R. FARWIG
Institut für Angewandte Mathematik, University of Bonn, D-5300 Bonn 1, West Germany

D. ZWICK
Department of Mathematics, University of Vermont, Burlington, VT 05405

Introduction. Let f be a real-valued function defined on an interval $[a, b]$ (that may be bounded or infinite) such that f' is continuous and piecewise continuously differentiable, and both f and f'' are bounded, and define

$$\|f\|_\infty = \sup\{|f(x)| : x \in [a, b]\}.^*$$

The *classical* (or *global*) Landau problem (see e.g. [1], [2], [7]) concerns finding explicit, sharp bounds on the numbers $\|f^{(k)}\|_\infty (k = 1, \dots, n-1)$, given bounds on $\|f\|_\infty$ and $\|f^{(n)}\|_\infty$. One variation ([3], [5], [8]) is to seek *pointwise* bounds; that is, to find sharp bounds on $|f^{(k)}(t)|$ $(k = 1, \dots, n-1)$ for each $t \in [a, b]$, given bounds on $\|f\|_\infty$ and $\|f^{(n)}\|_\infty$. In this paper we consider a variation of the pointwise problem in the case $n = 2$, one which has an interesting kinematic interpretation. For simplicity, assume for now that our interval is [0, 1].

Suppose that, for the functions described above, we additionally require that

$$\|f\|_\infty \le 1 \text{ on } [0,1] \quad \text{and} \quad -2m \le f''(s) \le 2M \text{ for all } s \in [0,1], \tag{1}$$

where $m, M > 0$. A function that satisfies these requirements will be called *admissible*. For fixed $t \in [0, 1]$ we seek the least number L such that

$$|f'(t)| \le L$$

for all such admissible functions f.

This problem has an interesting kinematic interpretation: A particle is allowed to move along

Reinhard Farwig received his *Diplom* (Master's degree) in 1980 from the University of Münster, West Germany, with a thesis on Functional Calculi and his Ph.D. in 1982 from the University of Bonn, having written his dissertation on a topic in Ordinary Differential Equations. He is currently working in the area of Mathematics of Computation with an emphasis on multivariate approximation. His interests include hiking, music, fine arts, and architecture.

Dan Zwick did his undergraduate work at the Hebrew University of Jerusalem and his graduate work at the University of Oregon, where he specialized in Approximation Theory. After a postdoctoral stay at the University of Bonn he accepted a position at Vermont, where he has been since 1982. Most of his work has involved generalized convexity in one form or the other. He enjoys travelling, collects art on a small scale, and sings in a local choral group.

*In their greatest generality, the results in this paper are true for

$$f \in W^2_\infty[a, b] = \{f : f' \text{ is absolutely continuous and } f'' \in L_\infty[a, b]\},$$

in which case one would use $\|f\|_\infty = \text{ess sup}\{|f(x)| : x \in [a, b]\}$.

a line in such a way that its distance from a fixed point cannot exceed one unit, and so that its acceleration *and deceleration* are bounded by $2M$ and $2m$, respectively (consider a car whose acceleration is determined by its engine power and whose deceleration is governed by its brakes and tires). We seek a strategy (i.e., an admissible function) that yields the largest possible velocity at a particular time t, given these constraints.

Let $t \in [0, 1]$ be fixed, and let α and β be numbers such that

$$0 \le t + \alpha \le t \le t + \beta \le 1. \tag{2}$$

We will show that an *optimal strategy*, one which achieves the maximal velocity L at time t, involves constant maximal acceleration starting at some time $t + \alpha$, and then constant maximal deceleration from time t to time $t + \beta$, where α and β are as in (2) (see Figs. 1–3). It makes sense that we also allow our particle the largest possible free run by setting

$$f(t + \alpha) = -1 \quad \text{and} \quad f(t + \beta) = 1.$$

If (as we will show) the maximal velocity can be reached in this fashion, then it is unimportant how the particle moves outside of the interval $[t + \alpha, t + \beta]$, provided the constraints (1) continue to be satisfied. It does turn out, however, that all optimal functions coincide in the interval $[t + \alpha, t + \beta]$.

Section 1. We now state the main theorem of this paper. We will restrict our attention to the interval [0, 1]; however the extension to an arbitrary closed and bounded interval is straightforward.

THEOREM 1. *Let f be admissible, i.e., defined on* [0, 1], *with f' continuous and piecewise continuously differentiable, satisfying* (1) *for some $m, M > 0$. Then there is a least function $L(t) > 0$, depending on m and M, such that $|f'(t)| \le L(t)$ for each $t \in [0, 1]$. $L(t)$ is continuously differentiable and is optimal in the sense that for each $t \in [0, 1]$ there is an admissible function $f_t(s)$ such that $f_t'(t) = L(t)$. Moreover, there is a subinterval containing t on which all admissible functions with this property coincide.*

Proof. Suppose that for α and β as in (2) we can find an admissible function f satisfying

$$f(t + \alpha) = -1, f(t + \beta) = 1, \tag{3}$$

$$f''(s) = 2M \quad \text{for} \quad t + \alpha < s < t \tag{4}$$
$$= -2m \quad \text{for} \quad t < s < t + \beta$$

and

$$f'(t + \alpha) = 0 \quad \text{if} \quad t + \alpha > 0, \tag{5}$$
$$f'(t + \beta) = 0 \quad \text{if} \quad t + \beta < 1.$$

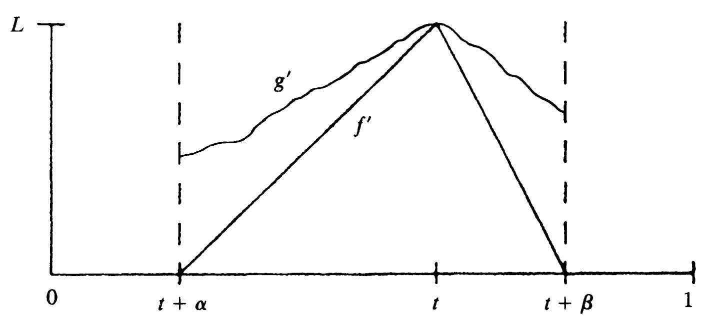


FIG. 1. Comparison of f' and g'.

If g denotes another admissible function with $g'(t) = f'(t)$, then the graph of g' must lie above the graph of f' (as shown in Fig. 1) since f' has the largest possible slope in $(t + \alpha, t)$ and the largest possible negative slope in $(t, t + \beta)$ (cf. [**6**, p. 422]). This implies that, unless $g \equiv f$ in $[t + \alpha, t + \beta]$,

$$g(t+\beta) = g(t+\alpha) + \int_{t+\alpha}^{t+\beta} g'(s)\,ds > g(t+\alpha) + \int_{t+\alpha}^{t+\beta} f'(s)\,ds$$
$$= g(t+\alpha) + f(t+\beta) - f(t+\alpha) = g(t+\alpha) + 2,$$

which is impossible if $\|g\|_\infty \le 1$. This shows that f has the largest possible value of $f'(t)$, i.e., is optimal, and that all optimal functions coincide in the interval $[t + \alpha, t + \beta]$.

Now for $s \in [t + \alpha, t + \beta]$ our optimal function f may be written as

$$f(s) = f(t) + L\cdot(s-t) + M\cdot(s-t)^2 - (m+M)(s-t)_+^2, \tag{6}$$

where $L = f'(t)$ and the *truncated power* $(s - t)_+^2$ is defined as $(s - t)^2$ if $s \ge t$, and is zero if $s \le t$. This *spline function* automatically satisfies the conditions on f''; the numbers L and $f(t)$, as well as α and β are to be determined from the remaining conditions.

We now address the problem of solving equations (3), (4) and (5). Due to the conditions (5), we distinguish the following four cases, depending on whether or not $t + \alpha$ is zero, and whether $t + \beta$ is one or less than one.

CASE 1. $t + \alpha > 0$ and $t + \beta < 1$, thus from (5) $f'(t + \alpha) = f'(t + \beta) = 0$ (see f' in Fig. 1). We get the equations

(7)
$$\begin{aligned}&(\mathrm{a})\ f(t+\alpha) = f(t) + L\alpha + M\alpha^2 = -1,\\ &(\mathrm{b})\ f(t+\beta) = f(t) + L\beta - m\beta^2 = 1,\\ &(\mathrm{c})\ f'(t+\alpha) = L + 2M\alpha = 0 \Rightarrow L = -2M\alpha,\end{aligned}$$

and

$$(\mathrm{d})\ f'(t+\beta) = L - 2m\beta = 0 \Rightarrow L = 2m\beta.$$

From (7c) and (7d) we have $2m\beta = L = -2M\alpha$, hence (7a) and (7b) give

$$f(t) = -1 - L\alpha - M\alpha^2 = -1 + M\alpha^2,$$

and similarly

$$f(t) = 1 - L\beta + m\beta^2 = 1 - m\beta^2,$$

which implies that

$$M\alpha^2 + m\beta^2 = 2.$$

Substituting $\beta = -M\alpha/m$, we get

$$\alpha^2 = \frac{m}{M}\left(\frac{2}{m+M}\right) \tag{8}$$

and

$$\beta^2 = \frac{M}{m}\left(\frac{2}{m+M}\right).$$

From (7c) and (8) we then have

$$L^2 = 4M^2\alpha^2 = \frac{8mM}{m+M},$$

from which

$$L = 2\left(\frac{2mM}{m+M}\right)^{1/2}.$$

If we define

$$\gamma = \left(\frac{2mM}{m+M}\right)^{1/2},$$

then we may write

$$\alpha = -\gamma/M, \quad \beta = \gamma/m \quad \text{and} \quad L = 2\gamma.$$

Finally,

$$f(t) = 1 - m\beta^2 = 1 - \frac{2M}{m+M} = \frac{m-M}{m+M}.$$

These computations are valid provided $t + \alpha > 0$ and $t + \beta < 1$, that is if

$$-\alpha < t < 1 - \beta.$$

Now, $-\alpha < 1 - \beta$ when $\gamma/M < 1 - \gamma/m$, and this condition is fulfilled precisely when

$$\frac{1}{m} + \frac{1}{M} < \frac{1}{2}. \tag{9}$$

Thus for m and M satisfying (9), and for $\gamma/M < t < 1 - \gamma/m$, the optimal solution f in the interval $[t + \alpha, t + \beta]$ is given by

$$f(s) = \frac{m-M}{m+M} + 2\gamma(s-t) + M(s-t)^2 - (m+M)(s-t)_+^2. \tag{10}$$

Note that f will continue to satisfy (1) if it is extended as a constant outside the interval $[t + \alpha, t + \beta]$, or if it is extended periodically, giving a kind of *perfect spline* **[6]**.

CASE 2. $0 = t + \alpha < t + \beta < 1$, hence $f'(t + \beta) = 0$ and $f'(0) \geq 0$ (see Fig. 2).

In this case equations (7a, b, d) hold with $t + \alpha = 0$, and (7c) is replaced by the condition

$$f'(0) = L - 2Mt \geq 0,$$

i.e.,

$$t \leq \frac{L}{2M}. \tag{11}$$

We are thus led to the equations

$$L = 2m\beta,$$
$$f(t) = 1 - m\beta^2,$$

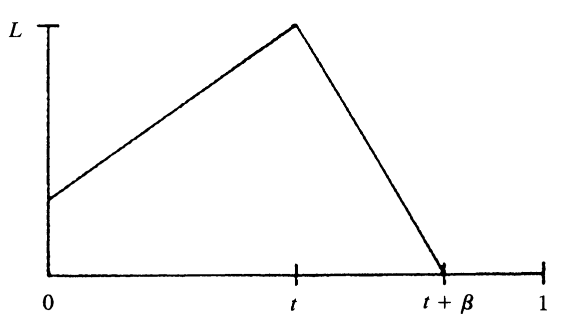


FIG. 2. $f'(s)$ in Case 2.

and by (7a)

$$\beta^2 + 2t\beta - \frac{Mt^2 + 2}{m} = 0,$$

whence

$$\beta = -t + \left(\frac{m + M}{m} t^2 + \frac{2}{m}\right)^{1/2}.$$

From the requirement that $t + \beta < 1$ we get

$$\frac{m + M}{m} t^2 + \frac{2}{m} < 1 \Rightarrow t^2 < \frac{m - 2}{m + M},$$

hence

$$m > 2 \quad \text{and} \quad t < \left(\frac{m - 2}{m + M}\right)^{1/2}.$$

From (11), after a straightforward calculation, we then get

$$t \leq \left\{\frac{m}{M}\left(\frac{2}{m + M}\right)\right\}^{1/2} = \frac{\gamma}{M}.$$

Now the inequality (9) of Case 1 can be shown to hold if and only if

$$\frac{\gamma}{M} < \left(\frac{m - 2}{m + M}\right)^{1/2};$$

thus Case 2 applies when $m > 2$ and

$$\text{(a)}\ \frac{1}{m} + \frac{1}{M} < \frac{1}{2} \quad \text{and} \quad t \leq \frac{\gamma}{M},$$

or

$$\text{(b)}\ \frac{1}{m} + \frac{1}{M} \geq \frac{1}{2} \quad \text{and} \quad t < \left(\frac{m - 2}{m + M}\right)^{1/2}.$$

CASE 3. $0 < t + \alpha < t + \beta = 1$, hence $f'(t + \alpha) = 0$ and $f'(1) \geq 0$.

This case is analogous to Case 2 (and can be derived from it by replacing t by $1 - t$ and β by $-\alpha$, and switching m and M), so we will content ourselves with giving the corresponding final equations.

We have

$$f(t) = -1 + M\alpha^2,$$
$$L = -2M\alpha$$

and

$$\alpha = (1 - t) - \left\{\left(\frac{m + M}{M}\right)(1 - t)^2 + \frac{2}{M}\right\}^{1/2}.$$

Case 3 applies when $M > 2$ and

$$\text{(a)}\ \frac{1}{m} + \frac{1}{M} < \frac{1}{2} \quad \text{and} \quad t \geq 1 - \frac{\gamma}{m},$$

or

$$\text{(b)}\ \frac{1}{m} + \frac{1}{M} \geq \frac{1}{2} \quad \text{and} \quad t > 1 - \left(\frac{M - 2}{m + M}\right)^{1/2}.$$

CASE 4. $t + \alpha = 0$ and $t + \beta = 1$, so that $f'(0) \geq 0$ and $f'(1) \geq 0$ (see Fig. 3).

The equations $f(0) = -1$ and $f(1) = 1$ yield

$$L = 2 + m(1-t)^2 + Mt^2$$

and

$$f(t) = -1 + Lt - Mt^2.$$

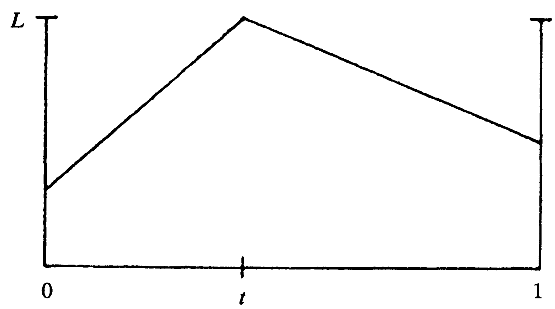


FIG. 3. $f'(s)$ in Case 4.

From $f'(0) \geq 0$ and $f'(1) \geq 0$ we get

$$(1-t)^2 \geq \frac{M-2}{m+M}$$

and

$$t^2 \geq \frac{m-2}{m+M}.$$

Thus Case 4 is valid in the following situations:

(a) $m \leq 2$ and $M \leq 2$,

(b) $m \leq 2$, $M > 2$ and $t \leq 1 - \left(\dfrac{M-2}{m+M}\right)^{1/2}$,

(c) $m > 2$, $M \leq 2$ and $t \geq \left(\dfrac{m-2}{m+M}\right)^{1/2}$,

or

(d) $m > 2$, $M > 2$ and $\left(\dfrac{m-2}{m+M}\right)^{1/2} \leq t \leq 1 - \left(\dfrac{M-2}{m+M}\right)^{1/2}$.

In order for (d) to make sense we need

$$\left(\frac{m-2}{m+M}\right)^{1/2} \leq 1 - \left(\frac{M-2}{m+M}\right)^{1/2},$$

which can be shown to hold precisely when

$$\frac{1}{m} + \frac{1}{M} \geq \frac{1}{2}. \tag{12}$$

Since at least one of m and M does not exceed 2 in each of (a)–(c), inequality (12) must also hold in these cases.

We now define $L(t)$ to be the value of L computed in each of the Cases 1–4. Thus, for fixed m and M,

$$
\begin{aligned}
L(t) &= 2\gamma && \text{if Case 1 applies,}\\
&= 2m\left[-t + \left(\frac{m+M}{m}t^2 + \frac{2}{m}\right)^{1/2}\right] && \text{if Case 2 applies,}\\
&= 2M\left[-(1-t) + \left(\frac{m+M}{M}(1-t)^2 + \frac{2}{M}\right)^{1/2}\right] && \text{if Case 3 applies,}\\
&= 2 + m(1-t)^2 + Mt^2 && \text{if Case 4 applies.}
\end{aligned}
$$

For completeness, and in order to show that $L(t)$ is continuously differentiable, we summarize the above analysis concerning the applicability of Cases 1–4: If $\gamma > 2$ (so that $m > 2$ and $M > 2$), then Case 1 applies if $t \in (\gamma/M, 1-\gamma/m)$; Case 2 applies if $t \in [0, \gamma/M]$, and Case 3 applies if $t \in [1-\gamma/m, 1]$. If $\gamma \le 2$, then Case 2 applies if $m > 2$ and $t < \left(\frac{m-2}{m+M}\right)^{1/2}$; Case 3 applies if $M > 2$ and $t > 1 - \left(\frac{M-2}{m+M}\right)^{1/2}$; otherwise Case 4 applies.

It is now a straightforward task, which we leave to the reader, to verify that $L(t)$ is continuously differentiable. ■

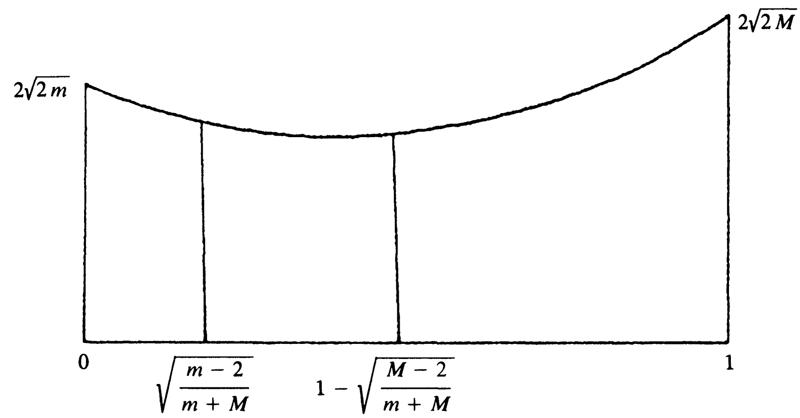


FIG. 4. $L(t)$ for $2 < m < M < 4$.

Fig. 4 shows the function $L(t)$ for a certain choice of m and M. The vertical lines indicate the transition from one case to another.

Note that as t approaches 1 the influence of M on $L(t)$ increases while the influence of m decreases, and vice versa as t approaches 0. This can be interpreted kinematically; for example the greatest velocity is achieved at $t = 1$ by constant maximal acceleration $2M$ in a certain time interval $[1+\alpha, 1]$, $\alpha < 0$, and thus $L(1)$ is independent of m (see Fig. 4). A similar result holds for $t = 0$, where $L(0)$ depends only on m.

When $m = M$, Theorem 1 recovers the pointwise results of [3], which were obtained by other means, and where due to symmetry only $t \in [0, \frac{1}{2}]$ was considered. Fig. 5 shows an optimal solution $f_t(s)$ for a typical choice of $m = M$. We observe that $L'(t)$ is nondecreasing in each instance and thus $L(t)$ is convex. Since it is (necessarily) symmetric with respect to $t = \frac{1}{2}$, we must have $L(0) = L(1) = \max\{L(t)\colon t \in [0,1]\}$, so that $L(0)$ solves the global Landau problem. Indeed, for $m = M$, $L(0)$ yields the same uniform bounds on $|f'(t)|$, for $t \in [0,1]$, calculated in [2].

Section 2. For the intervals $[0, \infty]$ and $(-\infty, \infty)$ the situation is simpler than for $[0,1]$. If m and M are positive, then the only possibility, in the case of $(-\infty, \infty)$, is that points $t+\alpha$ and $t+\beta$ exist in which f' vanishes. Thus Case 1 must apply (with no restrictions). Similarly, for the interval $[0, \infty]$ either Case 1 or Case 2 must apply, depending of course on m, M and t. Specifically, our previous calculations show that Case 2 is valid when $t < \gamma/M$.

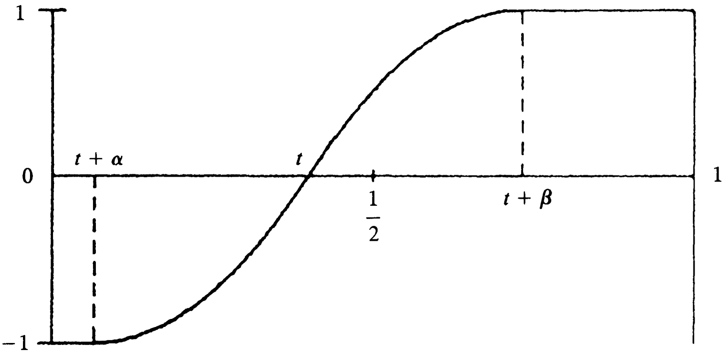


FIG. 5. An optimal function corresponding to $m = M > 4$ and $t \in (1/\sqrt{m}, 1/2)$.

REMARK. The idea of using asymmetric bounds for this type of problem apparently originated with L. Hörmander [**4**].

References

1. A. S. Cavaretta, Jr., An elementary proof of Kolmogorov's Theorem, this MONTHLY, 81 (1974) 480–486.

2. C. K. Chui and P. W. Smith, A note on Landau's Problem for bounded intervals, this MONTHLY, 82 (1975) 927–929.

3. R. Farwig and D. Zwick, The Pointwise Landau Problem for $n = 2$, University of Bonn, Sonderforschungsbereich 72 Preprint Series No. 613 (1983).

4. L. Hörmander, A new proof and a generalization of an inequality of Bohr, Math. Scand., 2 (1954) 33–45.

5. H. Kallioniemi, On bounds for the derivatives of a complex valued function on a compact interval, Math. Scand., 39 (1976) 295–314.

6. S. Karlin, Oscillatory Perfect Splines and Related Extremal Problems, in "Studies in Spline Functions and Approximation Theory", Editors S. Karlin, C. A. Micchelli, A. Pinkus, I. J. Schoenberg, Academic Press, New York, 1976, pp. 371–460.

7. I. J. Schoenberg, The elementary cases of Landau's Problem of inequalities between derivatives, this MONTHLY, 80 (1973) 121–158.

8. H. G. ter Morsche, Interpolation and Extremal Properties of *L*-Spline Functions, Dissertation, Tech. Univ. Eindhoven, 1982.

A QUICK ROUTE TO SUMS OF POWERS

A. W. F. EDWARDS

Gonville and Caius College, Cambridge, England

It is a remarkable fact that the polynomial for the sum of the rth powers of the integers

$$\sum_{v=1}^{n} v^r = a_1 n + a_2 n^2 + \cdots + a_{r+1} n^{r+1} \tag{1}$$

may be expressed in terms of the first two sums

A. W. F. Edwards: I am Reader in Mathematical Biology in the University of Cambridge. An interest in the foundations of statistics (manifest in my 1972 book *Likelihood*, Cambridge University Press) has led to historical researches in the origins of combinatorial theory, culminating in a book *Pascal's Arithmetical Triangle* (Griffin, High Wycombe; in press), and thence to Bernoulli's *Ars conjectandi* and the work of Faulhaber.

In my spare time I am President of the C.U. Gliding Club and holder of the international Gold badge for gliding (as is my wife also). I have contributed to the theory of cross-country soaring, which is a fascinating exercise in applied mathematics.

$$\sum_{v=1}^{n} v = n(n+1)/2$$

and

$$\sum_{v=1}^{n} v^2 = n(n+1)(2n+1)/6,$$

a result ultimately traceable to the symmetry of the Bernoulli polynomials. The first example is, of course, the familiar yet striking relation

$$\sum_{v=1}^{n} v^3 = \left\{ \sum_{v=1}^{n} v \right\}^2;$$

the general result was proved by Jacobi (1834).

But perhaps even more striking than the result itself is the fact that it was known over two centuries before Jacobi's time, by the forgotten German mathematician Johann Faulhaber, in whose *Academia algebrae* (Augspurg, 1631) I found it in 1981 whilst pursuing a lead from James Bernoulli's *Ars conjectandi* (Basel, 1713). At the point where he introduces the polynomials (1) and gives a table of the coefficients (including, of course, the Bernoulli numbers, so-called by de Moivre), Bernoulli mentions the name of Faulhaber. It so happened that the one work of Faulhaber readily accessible to me was the copy of *Academia algebrae* belonging to Cambridge University, oddly enough once the property of Jacobi (though whether he acquired it before or after 1834 we cannot say).

Writing $\sum_{v=1}^{n} v^r$ as $\sum n^r$ for simplicity, we find that Faulhaber's polynomials are

$$\sum n^r (r \text{ even}) = \sum n^2 \cdot \left(b_1 + b_2 \sum n + b_3 \left(\sum n\right)^2 + \cdots b_{r/2} \left(\sum n\right)^{(r/2)-1} \right)$$

and

$$\text{(2)} \quad \sum n^r (r \text{ odd} \geqslant 3) = \left(\sum n\right)^2 \cdot \left(c_1 + c_2 \sum n + c_3 \left(\sum n\right)^2 + \cdots + c_{(r-1)/2} \left(\sum n\right)^{(r-3)/2} \right),$$

where of course the coefficients b_i and c_i differ with each r.

Faulhaber gave an algorithm for obtaining the coefficients c_i from the b_i for the preceding value of r, and a method for obtaining the b_i themselves. I have christened the forms (2) 'Faulhaber polynomials' (Edwards, 1982), and Schneider (1983) has given an account of his methods. The purpose of the present paper is to exhibit the matrix forms for the Faulhaber polynomials in a way analogous to the matrix forms for the polynomials (1) (Edwards, 1982). Apart from its intrinsic elegance this approach allows Faulhaber's algorithm for obtaining the c_i from the b_i to be easily understood.

Consider the expansion of $[x(x+1)]^r - [x(x-1)]^r$, and apply to it Pascal's method of writing the identity successively for $x = 1, 2, 3, \ldots n$ and summing, as suggested by Tits (1923). We obtain (in the Σ notation introduced above)

$$\text{(3)} \qquad [n(n+1)]^r = 2\left[r \sum n^{2r-1} + \binom{r}{3} \sum n^{2r-3} + \binom{r}{5} \sum n^{2r-5} + \cdots \right],$$

which may be written in matrix form with rows for $r = 2, 3, 4, \ldots,$

$$\text{(4)} \qquad \begin{pmatrix} [n(n+1)]^2 \\ [n(n+1)]^3 \\ [n(n+1)]^4 \\ [n(n+1)]^5 \\ \cdot \end{pmatrix} = 2 \begin{pmatrix} 2 & & & & \\ 1 & 3 & & 0 & \\ 0 & 4 & 4 & & \\ 0 & 1 & 10 & 5 & \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} \begin{pmatrix} \sum n^3 \\ \sum n^5 \\ \sum n^7 \\ \sum n^9 \\ \cdot \end{pmatrix}$$

in which each row of the matrix is the corresponding row of Pascal's triangle with every other coefficient omitted. Writing $u = n(n+1)$ and solving (4) for the sums of the odd powers, we have

$$\begin{pmatrix} \sum n^3 \\ \sum n^5 \\ \sum n^7 \\ \sum n^9 \\ \cdot \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 2 & & & & \\ 1 & 3 & & 0 & \\ 0 & 4 & 4 & & \\ 0 & 1 & 10 & 5 & \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}^{-1} \begin{pmatrix} u^2 \\ u^3 \\ u^4 \\ u^5 \\ \cdot \end{pmatrix}, \tag{5}$$

which is the complete solution for the 'odd' Faulhaber polynomials since $u = 2\sum n$. In particular, u^2 is a factor of every polynomial, this proving the form given in (2).

Call the matrix of (4) $\mathbf{F} = \{f_{ij}\}$, then

$$f_{ij} = \binom{i+1}{2(i-j)+1}, \tag{6}$$

or zero for all values of i and j which do not define binomial coefficients. The matrix

$$\mathbf{F}^{-1} = \begin{pmatrix} 1/2 & & & & \\ -1/6 & 1/3 & & 0 & \\ 1/6 & -1/3 & 1/4 & & \\ -3/10 & 3/5 & -1/2 & 1/5 & \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} \tag{7}$$

thus leads to all the coefficients of the odd Faulhaber polynomials, when due allowance is made for the factor 2 in $u = 2\sum n$.

Tits also treated the even polynomials, applying Pascal's method to the expansion of $x^r(x+1)^{r+1} - x^{r+1}(x-1)^r$ and using (3) to remove the odd powers. The result, written in matrix form, is

$$\begin{pmatrix} \sum n^2 \\ \sum n^4 \\ \sum n^6 \\ \sum n^8 \\ \cdot \end{pmatrix} = \frac{1}{2}(2n+1) \cdot \begin{pmatrix} 3 & & & & \\ 1 & 5 & & 0 & \\ 0 & 5 & 7 & & \\ 0 & 1 & 14 & 9 & \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}^{-1} \begin{pmatrix} u \\ u^2 \\ u^3 \\ u^4 \\ \cdot \end{pmatrix}. \tag{8}$$

Call the matrix of (8) $\mathbf{G} = \{g_{ij}\}$, then

$$g_{ij} = \binom{i+1}{2(i-j)+1} + \binom{i}{2(i-j)+1}, \tag{9}$$

undefined binomial coefficients again being replaced by zeros. It will be seen that

$$\mathbf{G} = \mathbf{F} + \begin{pmatrix} 1 & 0 & 0 & 0 & \cdot \\ 0 & & & & \\ 0 & & & \mathbf{F} & \\ 0 & & & & \\ \cdot & & & & \end{pmatrix}. \tag{10}$$

The matrix

$$\text{(11)} \qquad \mathbf{G}^{-1} = \begin{pmatrix} 1/3 & & & & \\ -1/15 & 1/5 & & 0 & \\ 1/21 & -1/7 & 1/7 & & \\ -1/15 & 1/5 & -2/9 & 1/9 & \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}$$

thus leads to all the coefficients of the even Faulhaber polynomials.

Faulhaber knew, in essence, how to obtain $\mathbf{F}^{-1}$ from $\mathbf{G}^{-1}$. We may discover his algorithm as follows.

Take $\mathbf{F}$ and divide the elements of each column by the diagonal element in that column and thus write

$$\text{(12)} \quad \mathbf{F} = \begin{pmatrix} 2 & & & & \\ 1 & 3 & & 0 & \\ 0 & 4 & 4 & & \\ 0 & 1 & 10 & 5 & \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} = \begin{pmatrix} 2 & & & & \\ & 3 & & 0 & \\ & & 4 & & \\ 0 & & & 5 & \\ & & & & \cdot \end{pmatrix} \begin{pmatrix} 1 & & & & \\ 1/3 & 1 & & 0 & \\ 0 & 1 & 1 & & \\ 0 & 1/5 & 2 & 1 & \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}.$$

Now take $\mathbf{G}$ and divide the elements of each row by the diagonal element in that row and thus write

$$\text{(13)} \quad \mathbf{G} = \begin{pmatrix} 3 & & & & \\ 1 & 5 & & 0 & \\ 0 & 5 & 7 & & \\ 0 & 1 & 14 & 9 & \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} = \begin{pmatrix} 1 & & & & \\ 1/3 & 1 & & 0 & \\ 0 & 1 & 1 & & \\ 0 & 1/5 & 2 & 1 & \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} \begin{pmatrix} 3 & & & & \\ & 5 & & 0 & \\ & & 7 & & \\ 0 & & & 9 & \\ & & & & \cdot \end{pmatrix}.$$

Comparing (12) and (13) we see that the derived matrices are identical, an identity which, when analysed using (6) and (9), rests on a simple (and uninteresting) identity involving binomial coefficients.

Now write

$$\mathbf{X} = \begin{pmatrix} 2 & & & & \\ & 3 & & 0 & \\ & & 4 & & \\ & 0 & & 5 & \\ & & & & \cdot \end{pmatrix}$$

and

$$\mathbf{Y} = \begin{pmatrix} 3 & & & & \\ & 5 & & 0 & \\ & & 7 & & \\ & 0 & & 9 & \\ & & & & \cdot \end{pmatrix}$$

and we have $\mathbf{X}^{-1}\mathbf{F} = \mathbf{G}\mathbf{Y}^{-1}$ or, inverting,

$$\mathbf{F}^{-1}\mathbf{X} = \mathbf{Y}\mathbf{G}^{-1}$$

and thus

$$\text{(14)} \qquad \mathbf{F}^{-1} = \mathbf{Y}\mathbf{G}^{-1}\mathbf{X}^{-1}.$$

It is easily seen that premultiplication of $\mathbf{G}^{-1}$ by $\mathbf{Y}$ and postmultiplication by $\mathbf{X}^{-1}$ amounts to

multiplying the rows of $\mathbf{G}^{-1}$ by $3,5,7,9,\ldots$, respectively, and dividing the columns by $2,3,4,5,\ldots$, respectively. Thus to obtain the jth coefficient in the ith row of $\mathbf{F}^{-1}$ we take the corresponding coefficient in $\mathbf{G}^{-1}$, multiply by $(2i+1)$ and divide by $(j+1)$.

For example, when $i=4$ we multiply the four coefficients in the fourth row of $\mathbf{G}^{-1}$,

$$-\frac{1}{15},\frac{1}{5},-\frac{2}{9},\frac{1}{9},$$

by

$$\frac{9}{2},\frac{9}{3},\frac{9}{4},\frac{9}{5},$$

to obtain the fourth row of $\mathbf{F}^{-1}$,

$$-\frac{3}{10},\frac{3}{5},-\frac{1}{2},\frac{1}{5}.$$

Thus the polynomial

$$\sum n^8=\frac{1}{2}(2n+1)\left(-\frac{1}{15}u+\frac{1}{5}u^2-\frac{2}{9}u^3+\frac{1}{9}u^4\right)$$

leads to

$$\sum n^9=\frac{1}{2}\left(-\frac{3}{10}u^2+\frac{3}{5}u^3-\frac{1}{2}u^4+\frac{1}{5}u^5\right).$$

Faulhaber's actual algorithm is different because we have worked with $u=2\Sigma n$ rather than Σn, but the difference is, of course, trivial.

References

A. W. F. Edwards, Sums of powers of integers, Math. Gaz., 66 (1982) 22–8.

J. Faulhaber, Academia algebrae, Augspurg, 1631.

C. G. J. Jacobi, De usu legitimo formulae summatoriae Maclaurinianae, J. Reine Angew. Math., 12 (1834) 263–72.

I. Schneider, Potenzsummenformeln im 17. Jahrhundert, Historia Math., 10 (1983) 286–96.

L. Tits, Sur la sommation des puissances numériques, Mathesis, 37 (1923) 353–5.

Added in proof: Of related interest is B. L. Burrows and R. F. Talbot, Sums of powers of integers, this MONTHLY, 91 (1984) 394–403.

---

170. MISCELLANEA

There is now, and there always will be room in the world for good mathematicians of every grade of logical precision. It is almost equally important that the small band whose chief interest lies in accuracy and rigor should not make the mistake of despising the broader though less accurate work of the great mass of their colleagues, as that the latter should not attempt to shake themselves wholly free from the restraint the former would put upon them.

—Maxime Bôcher, Bull. Amer. Math. Soc., 10 (1904) 135.

A great teacher (one of von Neumann's) who bears the name of a great trigonometric theorem. (See p. 478.)

The assumption here that f is convex is not essential to the proof that $\lim_{n\to\infty} d_n$ exists. This result holds whenever f is positive and decreasing. Our method, however, allows us to take $f(x) = \log x$ in (1) and we obtain, after a short calculation, the existence of the limit in Stirling's formula

$$\alpha = \lim_{n\to\infty}\left(\frac{n!}{n^{n+1/2}e^{-n}}\right).$$

The proof that $\alpha = \sqrt{2\pi}$ uses Wallis' product (see, for example, [2, p. 363]).

This method of rearranging areas also gives good estimates for the speed of convergence in these limits. For example, if f is convex decreasing with $\lim_{x\to\infty} f(x) = 0$, we obtain rather easily that

$$\frac{1}{2}f(n+1) < d_n - \gamma_f < \frac{1}{2}f(n), \qquad n = 1,2,\dots. \tag{3}$$

Indeed, by considering only the sets $A_n, A_{n+1}, \dots,$ in Fig. 2 we see that

$$\sum_{k=n}^{\infty} a_k < \frac{1}{2}|f(n) - f(n+1)|, \qquad n = 1,2,\dots. \tag{4}$$

Now, for $N = n+1, n+2, \dots,$ we have

$$\sum_{k=n}^{N-1} a_k = \left|\sum_{k=n+1}^{N} f(k) + \frac{1}{2}(f(n) - f(N)) - \int_n^N f(x)\,dx\right|,$$

and so, if f is convex decreasing with $\lim_{x\to\infty} f(x) = 0$, then

$$0 < \gamma_f - d_n + \frac{1}{2}f(n) < \frac{1}{2}(f(n) - f(n+1)),$$

since

$$\gamma_f - d_n = \lim_{N\to\infty}\left(\sum_{k=n+1}^{N} f(k) - \int_n^N f(x)\,dx\right).$$

This proves (3).

In fact the lower estimate in (3) can be improved (see [1]) to $\frac{1}{2}f(n+\frac{1}{2})$. This improvement can also be demonstrated using the method of rearranging areas. We leave it as an amusing exercise for the reader to find a way of dissecting the sets $A_n, A_{n+1}, \dots,$ and rearranging the pieces so that they all lie, without overlapping, in a rectangle with sides $\frac{1}{2}$ and $|f(n) - f(n+\frac{1}{2})|$. This shows that the upper estimate in (4) can be improved to $\frac{1}{2}|f(n) - f(n+\frac{1}{2})|$, which gives the desired improvement of the lower estimate in (3).

References

1. R. P. Boas, Estimating remainders, Math. Mag., 51 (1978) 83–89.
2. R. Courant, Differential and Integral Calculus, Vol. 1, Blackie and Son, 1966.
3. I. Papadimitriou, A simple proof of the formula $\sum_{k=1}^{\infty} k^{-2} = \pi^2/6$, this Monthly, 80 (1973) 424–425.

ANSWER TO PHOTO ON PAGE 456

Lipót Fejér (1880–1959).

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description .of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics. The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

HOW SMALL CAN A SUM OF ROOTS OF UNITY BE?

GERALD MYERSON
*Department of Mathematics, SUNY at Buffalo, Buffalo, NY 14214**

Fix a positive integer N. How small in absolute value can a sum of five Nth roots of unity be? Well, zero; if N is a multiple of five, the roots can lie at the vertices of a regular pentagon, and sum to zero. If N is a multiple of six, three of the roots can lie at the vertices of an equilateral triangle, the other two directly opposite each other, and again the sum is zero. It is not hard to show that these are the only ways in which five Nth roots of unity can sum to zero; let us exclude them and talk only of non-zero sums. For fixed N, there are only finitely many ways of choosing five Nth roots (namely, $\binom{N+4}{5}$, if the roots are not required to be distinct), so the non-zero sums are strictly bounded away from zero. Let $f(N)$ be the least absolute value of a non-zero sum of five Nth roots of unity; we seek to estimate $f(N)$.

With a little algebraic number theory, we show that $f(N) > 5^{-N}$, as follows; any sum of five Nth roots of unity is an algebraic integer, so the product of all of its conjugates is at least one in absolute value. There are fewer than N of these conjugates, none bigger than five in absolute value, so the smallest conjugate exceeds 5^{-N} in absolute value.

We suspect that no non-zero sum of five Nth roots of unity can be anywhere near as small as 5^{-N}, and that $f(N) > c_5 N^{-2}$, or perhaps $f(N) > c_5 N^{-3}$, for some positive constant c_5, would be closer to the truth. The reason for our suspicion is related to the answer to the question: why five?

Let $f(k, N)$ be the least absolute value of a non-zero sum of k Nth roots of unity. If $k = 2$, it is clear that the minimal non-zero sum is achieved when the two roots are as nearly directly opposite as possible. A short calculation shows that

$$f(2, N) = 2\sin\frac{\pi}{N} \approx 2\pi N^{-1} \quad \text{if } N \text{ is even,}$$

$$f(2, N) = 2\sin\frac{\pi}{2N} \approx \pi N^{-1} \quad \text{if } N \text{ is odd.}$$

If $k = 3$, we assume, without loss of generality, that one of the roots of unity is 1. The other two must have imaginary parts opposite in sign and nearly equal in magnitude to encourage cancellation; from the geometry of the unit circle, their real parts are also nearly equal in magnitude. Therefore the real parts must be near minus one-half, and the three roots are near the vertices of an equilateral triangle. Now a simple calculation shows that the sum, assumed non-zero, cannot be smaller than $c_3 N^{-1}$ for some constant c_3. As N increases, we have

$$f(3, n) \sim \frac{2\pi\sqrt{3}}{3} N^{-1} \quad \text{for } N \text{ not divisible by 3,}$$

$$f(3, N) \sim 2\pi\sqrt{3}\, N^{-1} \quad \text{for } N \text{ divisible by 3.}$$

*Current address: Department of Mathematics, University of Texas, Austin, TX 78712.

The case $k = 4$ can also be settled by geometric considerations. We may assume that two of the roots are complex conjugate; to create maximal cancellation, the other two must be close to the negatives of the first two. Thus, the four roots lie near the vertices of a rectangle. Of all possible rectangles, the narrowest (*not* the squarest) turns out to be the best. The optimal configuration is to take two roots equal, the other two as close as possible to the negative of the first two, one on either side. For example, if $N = 2n + 1$ is odd, let the roots be 1, 1, $\exp(2\pi in/N)$, and $\exp(-2\pi in/N)$. We find

$$f(4, N) \sim \pi^2 N^{-2} \quad \text{if } N \text{ is odd,}$$
$$f(4, N) \sim 4\pi^2 N^{-2} \quad \text{if } N \text{ is even.}$$

The case $k = 5$ is the first we cannot solve by geometric considerations. We may assume that the first root is 1; once the second and third roots are chosen, the choice of fourth and fifth which minimizes the sum is easily determined. But the two degrees of freedom permitted in choosing the second and third seem to torpedo arguments from geometry.

We turn to the question of constructing, for given k and N, a small sum of k Nth roots of unity. We generally think of N as being large compared to k. Choosing roots near the vertices of the regular k-gon will never result in a non-zero sum smaller in magnitude than $c_k N^{-1}$ for some constant c_k; nevertheless, we know of no general construction better than this. For some special cases, we can do considerably better.

Consider the equation $1^r + 4^r + 4^r = 2^r + 2^r + 5^r$. This is known as a **multigrade** equation; it holds for $r = 0$, 1, and 2. It follows from the power series expansion of e^x that

$$|\theta + 2\theta^4 - 2\theta^2 - \theta^5| \sim 16\pi^3 N^{-3},$$

where $\theta = \exp(2\pi i/N)$. Thus, if N is even, so that $-\theta^2$ and $-\theta^5$ are Nth roots of unity, then $f(6, N) < c_6 N^{-3}$, for some constant c_6.

More generally, if there exist integers $a_0, \dots, a_s$ and $b_0, \dots, b_s$ such that

$$a_0^r + \cdots + a_s^r = b_0^r + \cdots + b_s^r \quad \text{for } r = 0, 1, \dots, l,$$

then it follows that for even N we have $f(2s + 2, N) < cN^{-(l+1)}$, where c depends only on s. Given l, how small can s be taken? It is easy to show that $s < l$ is impossible. It is conjectured [**7**] that s can always be taken equal to l, but this conjecture has only been verified for $l \leqslant 9$ [**2**]. Thus we know that, for N even, we have $f(20, N) < c_{20} N^{-10}$. It is known [**7**] that we can take $s \leqslant \frac{1}{2}(l^2 + 4)$; it follows that, for even N, and $k > 5$, $f(k, N) < c_k N^{-[\sqrt{k-6}\,]-1}$.

Another approach to constructing small sums is the following, used in [**3**]. Let p be a prime number. There are $\binom{p+k-1}{k}$ sums of k pth roots of unity. From Galois theory we know that if $k < p$, then no two of these sums are equal. Moreover, they all lie on or inside the circle $|z| = k$, so some pair of sums must be quite close. Just *how* close—that is, just how far apart one can spread a given number of points in a circle—is itself an unsolved problem. Early work was done by Pirl [**6**], and an up-to-date discussion is in [**4**]. Asymptotically, the sums cannot be spread farther apart than the points of the triangular lattice [**1**], so two of them must be within a distance

$$\left(\sqrt{2\pi/\sqrt{3}} + \varepsilon\right) k\sqrt{k!}\, p^{-k/2}$$

of each other (the special nature of the points in our problem, e.g., their p-fold rotational symmetry, may allow for some improvement here). If $N = 2p$, the difference of these two sums is a sum of at most $2k$ Nth roots of unity. Thus, given an integer N which is twice a prime, and given $k < N/2$, there exists $k' < 2k$ such that

$$f(k', N) \leqslant 2k 2^{k/2}\sqrt{k!}\, N^{-k/2}$$

(we have been generous with ε here). For large k this beats any bound that can now be proved using multigrades, but not the bound that would result from the conjecture on multigrades mentioned above.

With some care, the method of [3] produces small sums for any even N. Find a set of k Nth roots linearly independent over the rationals—this can be done for $k < \phi(N)$, where $\phi(N)$ is Euler's totient function. Now consider sums over subsets as before.

Is $f(k, N)$ monotone in N? For $k = 2, 3$, and 4, results given earlier in this paper show that, if $N_1 \equiv N_2 \pmod k$, and $N_1 > N_2$, then $f(k, N_1) < f(k, N_2)$. It is not clear whether this is true for any other value of k.

Is $f(k, N)$ monotone in k? For $N = 1, 2, 3, 4$ and 6 it is easy to see that $f(k, N) > 1$ for all k. For any other value of N, there is a sequence of values of k such that $f(k, N)$ approaches zero. Find a sum of Nth roots of unity which is less than 1 in magnitude, and consider its powers. Of course, this is very far from showing that, for fixed N, $f(k, N)$ is monotone; for all we know there are infinitely many N for which $f(5, N) > f(4, N)$. H. W. Lenstra has pointed out that there is monotonicity in k if N is divisible by six, since in that case

$$\omega = \omega \exp(2\pi i/6) + \omega \exp(-2\pi i/6)$$

expresses any Nth root as a sum of two Nth roots.

Given a non-zero sum α of k Nth roots of unity, let $L(\alpha)$ be the absolute value of the product of those conjugates of α which lie inside the unit circle (for those who know about such things, $L(\alpha)M(\alpha) = |N(\alpha)|$, where M is the measure and N is the norm). We close this collection of problems by asking

How small can $L(\alpha)$ be?

The argument from the second paragraph of this paper shows that $L(\alpha) > k^{-N}$; we ask whether there exist constants c_k and d_k depending only on k such that $L(\alpha) > c_k N^{-d_k}$. Any estimate of the form $\max_\alpha |\ln L(\alpha)| = o(N)$ for fixed k as N goes to infinity would have implications for problems in cyclotomy [5].

References

1. L. Fejes Tóth, Lagerungen in der Ebene, auf der Kugel und in Raum, Springer, 1953; MR 15, 248.
2. Albert Gloden, Mehrgradige Gleichungen, Noordhoff, Groningen, 1944; MR 8, 441f.
3. R. L. Graham and N. J. A. Sloane, Anti-Hadamard matrices, Linear Algebra Appl., 62 (1984) 113–137.
4. Willy Moser and Janos Pach, Research Problems in Discrete Geometry, 1984.
5. Gerald Myerson, A combinatorial problem in finite fields, II, Quart. J. Math. Oxford Ser. (2) 31 (1980) 219–231; MR 81i: 05014.
6. U. Pirl, Der Mindestabstand von n in der Einheits-kreisscheibe gelegenen Punkten, Math. Nachr., 40 (1969) 111–124; MR 40, 6379.
7. E. Maitland Wright, On Tarry's problem (I), Quart. J. Math. Oxford Ser. (1) 6 (1935) 261–267; Zbl. 13, 199a.

NOTES

Edited by Sabra S. Anderson, Sheldon Axler, and J. Arthur Seebach, Jr.
For instructions about submitting Notes for publication in this department see the inside front cover.

AUTOMORPHISM GROUPS

Gary L. Walls
Department of Mathematics, University of Southern Mississippi, Hattiesburg, MS 39406

Standard beginning algebra texts usually give the student a brief introduction to the notion of the automorphism group of a group. They then require the student to compute the automorphism

groups of a few groups of small order. In particular, students are often asked to show that $\mathrm{Aut}(D_6) \cong D_6$ and $\mathrm{Aut}(D_8) \cong D_8$. These examples can lead the student to some interesting conjectures about the structure of the automorphism groups of similar groups. This paper is an investigation (which could be done by a student) of some of these conjectures.

In this paper we let Z_n denote the cyclic group of order n and D_{2n} the dihedral group of order $2n$. The dihedral group consists of the symmetries of the regular n-gon.

Next we will review (very rapidly) the notion of a semi-direct product. Given two groups A, B, and a homomorphism θ: $B \to \mathrm{Aut}(A)$, the semi-direct product gives a group structure to the set $G = A \times B$ in which the multiplication is "twisted" by θ. The basic multiplication rule is

$$(a_1, b_1)(a_2, b_2) = (a_1 \cdot \theta(b_1)(a_2), b_1 b_2).$$

The reader who is unfamiliar with semi-direct products might show that G under this multiplication does form a group. In particular the student should at least verify the associative law. We will denote the semi-direct product of A by B with θ: $B \to \mathrm{Aut}(A)$ by $G = [A]_\theta B$ where θ may be omitted if it is clear from the context.

It is sometimes helpful when studying the semi-direct product $G = [A]_\theta B$ to make some "natural" identifications of A and B with subgroups of G. We identify an element $a \in A$ with $(a, 1)$ and $b \in B$ with $(1, b)$. With these identifications we can see that the elements of G can all be written uniquely in the form ab with $a \in A$ and $b \in B$. Multiplication of elements from A and B is determined by the relation $bab^{-1} = \theta(b)(a)$. In other words "conjugation by b" of elements of A is the same as applying the automorphism associated with b. Now the mystery of the multiplication for G can be cleared up. The idea is to write the element $(a_1 b_1)(a_2 b_2)$ in the correct form (an element from A times an element from B). But this is easy:

$$(a_1 b_1)(a_2 b_2) = \left(a_1\left(b_1 a_2 b_1^{-1}\right)\right)(b_1 b_2) = \left(a_1 \theta(b_1)(a_2)\right)(b_1 b_2)$$

as before (after the identifications).

From the above identifications we see that we can write $G = AB$ with $A \lhd G$ and $A \cap B = 1$. In particular we see that

$$G/B = AB/B \cong A/A \cap B = A.$$

It is in this sense that we say that a semi-direct product of A by B is an extension of A by B. Furthermore, it is a simple exercise to show that it is precisely these conditions which determine when a group is (isomorphic to) a semi-direct product. That is, if A, B are subgroups of G such that $A \lhd G$, $G = AB$, and $A \cap B = 1$, then there exists a θ: $B \to \mathrm{Aut}(A)$ such that $G \cong [A]_\theta B$. The mapping θ turns out to be the obvious map (seen above), $\theta(b)(a) = bab^{-1}$.

From this and our discussion of the dihedral group we can see that $D_{2n} \cong [\langle x \rangle]_\theta \langle y \rangle$, where θ: $\langle y \rangle \to \mathrm{Aut}(\langle x \rangle)$ is given by $\theta(y)(x) = x^{-1}$ (as this is yxy^{-1}).

We remark that if θ: $B \to \mathrm{Aut}(A)$ has the property that for all $b \in B$, $\theta(b) = I_A$, then $G = [A]_\theta B$ is isomorphic (actually equal) to the direct product of A and B. Also, if $B = \mathrm{Aut}(A)$ and θ is the identity map, then we call $G = [A]_I(\mathrm{Aut}(A))$ the holomorph of A and write $\mathrm{Hol}(A)$. The reader could probably show that $\mathrm{Hol}(Z_3) \cong D_6$ and $\mathrm{Hol}(Z_4) \cong D_8$.

In the beginning we were interested in the automorphism groups of dihedral groups. The next result gives us an interesting answer.

THEOREM A. *Let* $G = D_{2n}$. *Then* $\mathrm{Aut}\, G \simeq \mathrm{Hol}(Z_n)$.

Proof. The group D_{2n} has presentation $\langle x, y | x^n, y^2, yxyx \rangle$. Using properties of presentations we know that a function α: $\{x, y\} \to D_{2n}$ can be extended to an automorphism of D_{2n} if and only if $\alpha(x)$, $\alpha(y)$ satisfy the same relations as x and y, and $\alpha(x)$, $\alpha(y)$ generate D_{2n}. That is, we need

$$(\alpha(x))^n = 1, \quad (\alpha(y))^2 = 1, \quad \alpha(y)\alpha(x)\alpha(y)\alpha(x) = 1, \quad \text{and} \quad \langle \alpha(x), \alpha(y) \rangle = D_{2n}.$$

Easy calculations show that all elements of D_{2n} are of the form x^i or $x^i y$ and that all elements of

the form $x^i y$ have order 2.

Since an automorphism must preserve order we see that we must have $\alpha(x) = x^i$ and $\alpha(y) = x^j y$ for some i, j satisfying $(i, n) = 1$ and $0 \leqslant j \leqslant n-1$. We cannot let $\alpha(y)$ be a power of x since we require that $D_{2n} = \langle \alpha(x), \alpha(y) \rangle$. But it is easy to check that all the allowable choices for i, j do give rise to automorphisms. Thus, $|\mathrm{Aut}(D_{2n})| = \phi(n) \cdot n$. Here $\phi(n)$ is the Euler ϕ-function. We wish to determine the structure of $A = \mathrm{Aut}(D_{2n})$. Let θ be defined by $\theta(x) = x$, $\theta(y) = xy$. It is easy to see that $\langle \theta \rangle \lhd A$. Also, let γ_i be defined by $\gamma_i(x) = x^i$, $\gamma_i(y) = y$. Then $B = \{\gamma_i | (i, n) = 1\}$ is a group and $B \cong \mathrm{Aut}(Z_n)$. Clearly, we have $A = \langle \theta \rangle B$. It follows that A is isomorphic to a semi-direct product of $\langle \theta \rangle$ and B. As B is the automorphism group of $\langle \theta \rangle$ we see that $\mathrm{Aut}(D_{2n})$ is isomorphic to $\mathrm{Hol}(Z_n)$. The actual construction of the isomorphism is left to the reader.

To check with our original results, we let $n = 3$ and 4. When $n = 3$ we see that

$$\mathrm{Aut}(D_6) \cong \mathrm{Hol}(Z_3) = [Z_3]Z_2 \cong D_6 \quad \text{and} \quad \mathrm{Aut}(D_8) \cong \mathrm{Hol}(Z_4) = [Z_4]Z_2.$$

It follows that these are the only two values of n for which we get $\mathrm{Aut}(D_{2n}) \cong D_{2n}$.

We include a similar theorem which we think is interesting.

THEOREM B. *Suppose* $G = [Z_n]_\theta Z_m$ *and* $Z(G) = 1$. *Then*, $\mathrm{Aut}\, G \cong \mathrm{Hol}(Z_n)$.

Proof. We will think of G as in the form $G = \langle x \rangle \langle y \rangle$ with $\langle x \rangle \lhd G$. Hence,

$$x^n = 1, \qquad y^m = 1, \qquad yxy^{-1} = x^k.$$

First, we note that as $Z(G) = 1$, we must have $\ker \theta = 1$. For if $y^l \in \ker \theta$, then

$$y^l x y^{-l} = \theta(y^l)(x) = I(x) = x,$$

so y^l commutes with x and y. Hence, $y^l \in Z(G) = 1$.

Now we can identify G with a subgroup of $\mathrm{Hol}(\langle x \rangle)$. We have the following sequence of subgroups (after the identifications);

$$\langle x \rangle \subseteq G \subseteq \mathrm{Hol}(\langle x \rangle).$$

Recall that $\langle x \rangle \lhd \mathrm{Hol}(\langle x \rangle)$ and $\mathrm{Hol}(\langle x \rangle)/\langle x \rangle \simeq \mathrm{Aut}(\langle x \rangle)$ is abelian. (The automorphism group of all cyclic groups are easily determined (see [**3**], page 117). In any case it is easily seen to be abelian.) It follows that $\langle x \rangle$ contains the commutator subgroup of $\mathrm{Hol}(\langle x \rangle)$. Hence, $G \lhd \mathrm{Hol}(\langle x \rangle)$.

This enables us to define a homorphism T: $\mathrm{Hol}(\langle x \rangle) \to \mathrm{Aut}(G)$ by

$$T(z)(g) = zgz^{-1}$$

for $z \in \mathrm{Hol}(\langle x \rangle)$ and $g \in G$. Clearly, T is a homomorphism. Suppose $x^i l \in \ker T$ with $l \in \mathrm{Aut}(\langle x \rangle)$. Then $(x^i l)(x)(x^i l)^{-1} = x$ implies that $lxl^{-1} = x$. But in the holomorph lxl^{-1} is the same as applying the automorphism l to x. Hence $l(x) = x$ implies that $l = 1$. Now $x^i \in \ker T$ implies $x^i \in Z(G) = 1$.

Hence, T is one-to-one. The proof will be completed by showing that $|\mathrm{Aut}\, G| \leqslant |\mathrm{Hol}(\langle x \rangle)|$ as this will imply that T is onto and, hence, an isomorphism. The proof proceeds in the following steps.

(1) As $Z(G) = 1$, $(n, k-1) = 1$.

If $(n, k-1) \neq 1$. Then $x^{\frac{n}{(n,k-1)}} \neq 1$ and

$$yx^{\frac{n}{(n,k-1)}}y^{-1} = x^{\frac{nk}{(n,k-1)}} = x^{\frac{n(k-1)}{(n,k-1)}} \cdot x^{\frac{n}{(n,k-1)}} = x^{\frac{n}{(n,k-1)}}.$$

So $x^{\frac{n}{(n,k-1)}} \in Z(G)$, a contradiction.

Let $\alpha \in \operatorname{Aut} G$. The following must be true.

(2) There exists i, $(i, n) = 1$, so that $\alpha(x) = x^i$. Suppose that $\alpha(x) = x^a y^b$ and $\alpha(y) = x^c y^d$. We must have

$$\alpha(y)\alpha(x)\alpha(y^{-1}) = \alpha(x^k).$$

It follows that

$$(x^c y^d)(x^a y^b)(x^c y^d)^{-1} = (x^a y^b)^k$$

and that (look in the quotient group $G/\langle x\rangle$) $y^b = y^{bk}$, so $m|b(k-1)$.

But now $|x^a y^b| = |x| = n$, so as $(n, k-1) = 1$, we must have $|(x^a y^b)^{k-1}| = n$. However,

$$(x^a y^b)^{k-1} = x^{a(k^{(k-2)b} + \cdots + k^b + 1)} y^{b(k-1)} = x^{a((k^{(k-1)b} - 1)/(k^b - 1))}.$$

This implies

$$\left(a\left(\frac{k^{(k-1)b} - 1}{k^b - 1}\right), n\right) = 1,$$

contrary to the fact that $n|k^{(k-1)b} - 1$ (recall that $m|(k-1)b$ and $k^m \equiv 1 \pmod{n}$) and $n \nmid k^b - 1$ (as $b < m$ and $\ker\theta = 1$).

(3) There exists a j such that $\alpha(y) = x^j y$.

We have established that $\alpha(x) = x^i$, so suppose $\alpha(y) = x^a y^b$. Then, as above,

$$(x^a y^b) x^i (x^a y^b)^{-1} = x^{ki}, \quad \text{so} \quad y^b x^i y^{-b} = x^{ki}.$$

Hence, $x^{k^b i} = x^{ki}$. Thus, $n|(k^b - k)i$ and we must have $n|k^b - k$.

It follows that $y^{b-1} \in C(\langle x\rangle) = \langle x\rangle$. Hence, we must have $b = 1$.

Now we see that at most there are $\phi(n) \cdot n$ choices for α. Hence,

$$|\operatorname{Aut} G| \leqslant n \cdot \phi(n) = |\operatorname{Hol}(\langle x\rangle)|.$$

It follows that T is onto and, hence, is an isomorphism.

The problem of determining $\operatorname{Aut}([Z_n]Z_m)$ becomes more difficult in the case where the center in not necessarily trivial. We invite the reader to investigate the general case. We suggest he begin by detailed study of the automorphism groups of the following groups:

(1) $\langle x, y | x^{12}, y^2, yxyx^5\rangle$, (2) $\langle x, y | x^{12}, y^2, yxyx^7\rangle$ (3) $\langle x, y | x^{16}, y^4 yxyx^{13}\rangle$.

References

1. I. Herstein, Topics in Algebra, Wiley, New York, 1975.
2. L. C. Mathewson, On the group of isomorphisms of a certain extension of an abelian group, Trans. Amer. Math. Soc., 22 (1918) 331–340.
3. W. R. Scott, Group Theory, Prentice-Hall, Englewood Cliffs, NJ, 1964.

THE FACTORIZATION OF A SQUARE MATRIX INTO TWO SYMMETRIC MATRICES

A. J. Bosch

Department of Mathematics, University of Technology, Eindhoven, The Netherlands

1. Introduction. As we will see, every square matrix (real or complex) is a product of two symmetric matrices (real or complex, respectively). Although these results were already published by Frobenius in 1910 (see [**2**]), they are still not well known to mathematicians. I could not find them in modern textbooks on matrix theory or linear algebra. Consequently, these results and their proofs (see [**1**], [**4**], [**5**]) are not very accessible to mathematicians working in other fields. The

aim of this paper is to give elementary proofs. The basis of the proofs is the Jordan normal form of a matrix.

2. Notations.

A is a complex or real square matrix;
A^T is the transpose of A;
$\bar{A}$ is the conjugate of A;
$A \simeq D$ means A is similar to the matrix D: $A = BDB^{-1}$;

$J_k(\lambda)$ is the $k \times k$-matrix $\begin{bmatrix} \lambda & 1 & & \\ & \ddots & \ddots & 0 \\ 0 & & \ddots & 1 \\ & & & \lambda \end{bmatrix}$.

3. Preliminaries.

DEFINITION 1. A Jordan matrix J is a square matrix of the form

$$J = \begin{bmatrix} J_{k_1}(\lambda_1) & & & \\ & \ddots & & 0 \\ 0 & & \ddots & \\ & & & J_{k_r}(\lambda_r) \end{bmatrix};$$

the λ_i are not necessarily different.

PROPOSITION 1 (Jordan normal form of a matrix; for the proof see [3]). *Let A be a complex square matrix. Then $A = BJB^{-1}$ for some invertible matrix B and some Jordan matrix J.*

PROPOSITION 2. *Every Jordan matrix is a product of a real symmetric matrix and a complex symmetric matrix.*

Proof. Let $S_k = \begin{bmatrix} 0 & & 1 \\ & \cdot^{\cdot^{\cdot}} & \\ 1 & & 0 \end{bmatrix}$ and $C_k(\lambda) = \begin{bmatrix} 0 & & \lambda & 1 \\ & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \\ \lambda & 1 & & 0 \end{bmatrix}$. Both are symmetric; S_k is real. Then $J_k(\lambda) = S_k C_k(\lambda)$. If J is as in Definition 1, then $J = SC$, where $S = \begin{bmatrix} S_{k_1} & & \\ & \ddots & 0 \\ 0 & & S_{k_r} \end{bmatrix}$ and $C = \begin{bmatrix} C_{k_1}(\lambda_1) & & \\ & \ddots & 0 \\ 0 & & C_{k_r}(\lambda_r) \end{bmatrix}$. Both are symmetric; S is real. Note that S is nonsingular.

DEFINITION 2. A Jordan matrix J is called *balanced* if, when $J_k(\lambda)$ is in J, then $\overline{J_k(\lambda)}$ is also in J with the same multiplicity. This means that $J \simeq \bar{J}$, or equivalently $A \simeq \bar{A}$, where A and J are related as in Proposition 1.

COROLLARY. *If A is a real square matrix and $A = BJB^{-1}$ as in Proposition 1, then J is balanced.*

Proof. $A = \bar{A}$, hence $J \simeq \bar{J}$.

4. Two theorems on factorization.

THEOREM 1. *Every complex square matrix A is a product of two complex symmetric matrices.*

Proof. If $A = BJB^{-1}$ as in Proposition 1, then Proposition 2 gives

$$A = B(SC)B^{-1} = BSB^T(B^T)^{-1}CB^{-1} = [BSB^T]\left[(B^{-1})^T CB^{-1}\right].$$

BSB^T and $(B^{-1})^TCB^{-1}$ are symmetric. Note that the first factor BSB^T is nonsingular.

THEOREM 2. *Every real square matrix A is a product of two real symmetric matrices.*

Proof. Let $A = BJB^{-1}$ as in Proposition 1 and $J = SC$ as in Proposition 2. Partition the matrices B, J and S all in the same way with respect to the columns:

$$B = \left(B_{(1)}B_{(2)}B_{(3)}\right); \quad J = \begin{bmatrix} J_{(1)} & & 0 \\ & J_{(2)} & \\ 0 & & J_{(3)} \end{bmatrix}; \quad S = \begin{bmatrix} S_{(1)} & & 0 \\ & S_{(2)} & \\ 0 & & S_{(3)} \end{bmatrix}.$$

A is real, hence J is balanced (see Corollary). By permutation of the columns of B we can see to it that $J_{(2)} = \bar{J}_{(1)}$ (where the λ's in $J_{(1)}$ are all different from the λ's in $\bar{J}_{(1)}$) and that $J_{(3)}$ contains *all* real blocks of J (so $J_{(3)}$ is real and $S_{(2)} = \bar{S}_{(1)} = S_{(1)}$). Of course $J_{(1)}$ or $J_{(3)}$ can be "empty". $AB = BJ$ in partitioned form gives:

$$\left(AB_{(1)}\ AB_{(2)}\ AB_{(3)}\right) = \left(B_{(1)}J_{(1)}\ B_{(2)}\bar{J}_{(1)}\ B_{(3)}J_{(3)}\right).$$

$AB_{(1)} = B_{(1)}J_{(1)}$, hence $A\bar{B}_{(1)} = \bar{B}_{(1)}\bar{J}_{(1)}$; $AB_{(3)} = B_{(3)}J_{(3)}$, hence $A\bar{B}_{(3)} = \bar{B}_{(3)}J_{(3)}$ ($B_{(3)}$ is of full rank). This means that there exists a *real* matrix $D_{(3)}$ of full rank such that $AD_{(3)} = D_{(3)}J_{(3)}$. Define $D = (B_{(1)}\bar{B}_{(1)}D_{(3)})$; then $AD = DJ$. D is nonsingular: $B_{(1)}$ (so $\bar{B}_{(1)}$) and $D_{(3)}$ are of full rank. The columns of $B_{(1)}$ and $\bar{B}_{(1)}$, and the columns of $B_{(1)}(\bar{B}_{(1)})$ and $D_{(3)}$ are mutually *independent*, being (generalized) eigenvectors associated with *different* eigenvalues. Hence D is nonsingular. Now

$$A = DJD^{-1} = DSCD^{-1} = [DSD^T]\left[(D^{-1})^TCD^{-1}\right].$$

Both factors are symmetric and DSD^T is, as a product of nonsingular matrices, nonsingular. In partitioned form we have:

$$DSD^T = \left(D_{(1)}S_{(1)}D_{(1)}^T + \bar{D}_{(1)}S_{(1)}\bar{D}_{(1)}^T\right) + D_{(3)}S_{(3)}D_{(3)}^T$$

which is, as a sum of two real matrices, also real.

$(D^{-1})^TCD^{-1} = (DSD^T)^{-1}A$ is, as a product of real matrices, also real.

References

1. D. H. Carlson, On real eigenvalues of complex matrices, Pacific J. Math., 15 (1965) 1119.

2. G. Frobenius, Ueber die mit einer Matrix vertauschbaren Matrizen, Sitzungsber. Preuss. Akad. für Wiss. (1910) 3.

3. Ben Noble and J. W. Daniel, Applied Linear Algebra, Prentice-Hall, 1977.

4. Chi Song Wong, Characterization of products of symmetric matrices, Linear Algebra and Appl., 42 (1982) 243.

5. O. Taussky, The role of symmetric matrices in the study of general matrices, Linear Algebra and Appl., 5 (1972) 147.

HOW MANY ZEROS DOES A CONTINUOUS FUNCTION HAVE?

TOMÁS DOMÍNGUEZ BENAVIDES
Facultad de Matemáticas, c/Tarfia s/n, Sevilla-12, Spain

Let $I = [a, b]$ and let $\mathscr{C}_O(I)$ be the space of those continuous real-valued functions on I having at least one zero with the topology of uniform convergence. We denote by d the metric in $\mathscr{C}_0(I)$,

$$d(f, g) = \max\{|f(x) - g(x)|: x \in I\}.$$

Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T: Textbook	P: Professional Reading	1-4: Semesters
C: Computer Software	L: Undergraduate Library	** : Special Emphasis
S: Supplementary Reading	13: Grade Level	?? : Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the Monthly.

General, P*. Mathematical Sciences: A Unifying and Dynamic Resource. National Academy Pr, 1986, 35 pp, (P). First in an NRC series of concise surveys of research mathematics intended primarily for non-mathematicians who help set national research policy. Six vignettes (D-modules, computational complexity, conservation laws, Yang-Mills equations, operator algebras, survival analysis) illustrate the resonance of theory and applications in contemporary mathematics. Includes data as a sequel to the David Report showing continued under-funding of mathematics research. LAS

General, P. Lecture Notes in Mathematics-1111: Arbeitstagung Bonn 1984. Ed: F. Hirzebruch, J. Schwermer, S. Suter. Springer-Verlag, 1985, 481 pp, $29.20 (P). [ISBN: 0-387-15195-8] These proceedings of a meeting held at the Max-Planck-Institut für Mathematik in Bonn, June 15-22, 1984 contain a number of survey lectures and refinements of ad-hoc presentations on various areas of geometry--both algebraic and differential as well as other related mathematics, including algebraic number theory. JAS

General, P. Topics in Modern Mathematics: Petrovskii Seminar No. 5. Ed: O.A. Oleinik. Contemp. Soviet Math. Plenum, 1985, 342 pp, $69.50. [ISBN: 0-306-10980-8] English translations of papers in group theory, differential equations, and differential geometry. JAS

General, S*, L.** The William Lowell Putnam Mathematical Competitions, Problems and Solutions: 1965-1984. Ed: Gerald L. Alexanderson, Leonard F. Klosinski, Loren C. Larson. MAA, 1985, xii + 147 pp, $24. [ISBN: 0-88385-441-4] A corrected reprinting from the Monthly and Mathematics Magazine of problems and solutions to the last 20 Putnam exams, intended not as a scholarly sequel to the first Putnam volume covering 1938-1965, but as a handy reference for problem solvers and undergraduate problem groups. Includes a brief account by Herbert Robbins of the first Putnam exam, and lists of team and individual winners. LAS

General, P, L. Women and Minorities in Science and Engineering. Ed: Michael F. Crowley, Melissa J. Lane. NSF, 1986, xii + 182 pp, (P). A statistical report of employment, salaries, education, disciplines, career patterns and labor market indicators for women and minorities in science and engineering. Sample: the percentages of white, black and Asian students taking high school calculus are 8.3%, 3.6% and 19.4%, respectively. LAS

General, S*(13-14). A Collection of Problems. Ya. S. Bugrov, S.M. Nikolsky. Transl: Leonid Levant. Higher Math. MIR (US Distr: Imported Pub), 1984, 191 pp, $5.95. [ISBN: 0-8285-2896-9] Keyed to three previously published texts by the same authors covering calculus of one and several variables, differential equations, linear algebra and complex variables. Twelve-hundred problems--mostly routine, some challenging. Answers for all, hints for some. Worked-out examples. Good problem source at low cost. JK

Elementary. Florida College Level Academic Skills Test Handbook. Dauhrice Gibson, Janice McFatter. Prentice-Hall, 1985, ix + 421 pp, $15.95 (P). [ISBN: 0-13-322421-X-01] Totally skills-oriented review for computations section of the CLAST. Five question pre-test and post-test bracket. Short section on each skill consisting primarily of examples. Covers basic aspects of arithmetic, algebra, logic, geometry, probability, and statistics. Contains sample test. MW

Mathematics Appreciation, S*(13-18), P*, L*. The Ins and Outs of Peg Solitaire. John D. Beasley. Rec. in Math., V. 2. Oxford U Pr, 1985, xii + 275 pp, $16.95. [ISBN: 0-19-853203-2] The second in the publisher's Recreation in Mathematics Series. Details of the rules, notation, equipment and history of the 33-hole version of peg solitaire. Underlying theory is explored as regards solubility, fewest-moves solutions and otherwise good play. Numerous diagrams (over 500) and problems (over 200). Concludes with several chapters on other-shaped boards, other rules of play and suggestions for further research. Excellent (selected) bibliography, glossary and index. Authoritative and enjoyable. For the novice as well as the expert. JK

Precalculus, T(13: 1). Precalculus Mathematics, Third Edition. Daniel D. Benice. Prentice-Hall, 1986, x + 566 pp, $30.95 (P). [ISBN: 0-13-695503-7] Now in paperback at a 40% increase in price, this edition expands the treatment of inverse functions, polynomial and rational functions, geometric applications and mathematical induction; includes additional examples and exercises and encourages calculator usage (Second Edition, TR, August-September 1982). JNC

Precalculus, S(13).** Problem Book in High-School Mathematics. Ed: A.I. Prilepko. Transl: I.A. Aleksanova. MIR (US Distr: Imported Pub), 1985, 280 pp, $7.95. [ISBN: 0-8285-3042-4] These 2000-plus non-routine, but not especially difficult problems, all with answers and some with solutions, provide an excellent review of materials much of which was once high school mathematics in this country. Outstanding sections/chapters on graphing, trigonometry and solid geometry. According to the Preface, most of the problems have been used in entrance examinations to colleges in the USSR in recent years. Entrance exams? Recommended problem source for teachers of pre-calculus courses. JK

Precalculus, T(13: 1). College Algebra, Second Edition. Marshall D. Hestenes, Richard O. Hill, Jr. Prentice-Hall, 1986, x + 437 pp, $29.95. [ISBN: 0-13-140856-9-01] With logarithmic and exponential tables, this Second Edition is more flexible as regards the use of the calculator. Some restructuring, some new material, additional exercises. Emphasis on readability. Topics covered are the usual. Middle level of difficulty. Prerequisite is one and one-half years of high school algebra/geometry. For one-term courses. Good preparation for courses in calculus, finite mathematics, discrete mathematics and statistics. JK

Education, P. The Computer and the Child: A Montessori Approach. Peter G. Gebhardt-Seele. Comp. in Educ. Ser. Computer Science Pr, 1985, xi + 250 pp, $19.95 (P). [ISBN: 0-88175-013-1] A look at computer education from Montessori perspective. Includes rationale, specific computer applications (e.g., games, CAI, drill and practice, word processing, programming, problem solving), sample materials and presentations, and useable LOGO and BASIC programs. A refreshing look at the teaching and learning of computing. RD

Education, P. Methods of Evaluating College Remedial Mathematics Programs: Results of a National Study. Geoffrey Akst, Susan Remmer Ryzewic. Res. Mono. Ser. Rep., No. 10. Instructional Resource Center (CUNY), 1985, 107 pp, $2 (P). A report with 80 tables summarizing results of a nation-wide study of remedial mathematics programs which enroll one out of every three post-secondary students. The focus of the study is on means of evaluating such programs. Includes several thoughtful recommendations with accompanying rationale. LAS

Education, P, L. Second International Mathematics Study: Summary Report for the United States. F. Joe Crosswhite, et al. US National Coordinating Center (U. of Ill., Urbana-Champaign), 1985, xix + 139 pp, (P). The official revised U.S. summary of the Second International Mathematics Study, superceding the 1984 preliminary report (TR, February 1986). On most measures, U.S. students perform in the lowest quartile in this study. LAS

Education, P. Microcomputers in Education Conference: Literacy Plus. Ed: Ruth A. Camuse. Computer Science Pr, 1984, xi + 465 pp, $35. [ISBN: 0-88175-077-8] Proceedings from the fourth annual conference, March 1984, at Arizona State University. Sixty-five papers cover computer literacy: research reports, program suggestions and evaluations, and specific ideas for content-area classroom teachers, from home economics to special education. RD

Education, T(14-15: 1, 2). Mathematics, An Informal Approach, Second Edition. Albert B. Bennett, Jr., Leonard T. Nelson. Allyn & Bacon, 1985, xvi + 711 pp, $32.85. [ISBN: 0-205-08305-6] Elementary mathematics (sets, arithmetic, fractions and decimals, geometry, and probability and statistics) for education majors, with an emphasis on problem solving (Pólya style). Pleasant browsing: many interesting examples and tidbits of information. Plenty of exercises. (First Edition, TR, November 1979.) BC

Education, S(15-16), P. Personal Computers for Education. Alfred Bork. Harper & Row, 1985, x + 179 pp, $19.95 (P). [ISBN: 0-06-040866-9] The author considers computers from a pedagogical perspective rather than from the typical technical approach. Three divisions in the book consider 1) using the computer in education, 2) the future of computers in education, and 3) a hodge-podge of issues for the educator to consider (hardware, psychological background, classroom use, production of computer-based learning materials). Stress is on looking toward the future--what is in use now will be inadequate very soon. RD

History, S*, P, L.** William Rowan Hamilton: Portrait of a Prodigy. Sean O'Donnell. Boole Pr, 1983, xvi + 224 pp, $24.95. [ISBN: 0-906783-06-2] A fascinating, well-documented, yet extraordinarily lucid account of the life and accomplishments of Hamilton, Ireland's greatest scientist. Focussing first on Hamilton's early development as a child prodigy, science journalist O'Donnell traces the growth of genius through Hamilton's career in mathematics, astronomy, poetry, and philosophy. Unfailingly attentive to detail, O'Donnell documents and analyzes both original sources and other biographical accounts to insure an accurate portrait. LAS

History, S, P*, L.** George Boole: His Life and Work. Desmond MacHale. Boole Pr, 1985, xiii + 304 pp, $24.95. [ISBN: 0-906783-05-4] The first full-length biography of Boole, blending tales of the child prodigy, social idealist, mathematical theorist, practical scientist, poet, husband, and father. Mathematician-biographer MacHale discusses Boole's contributions to mathematics with superb polish and professionalism, relating his work both to the context of his life and to its implications for future applications. LAS

History, S*(15-17), P*. Pappus of Alexandria, Book 7 of the Collection. Ed: Alexander Jones. Springer-Verlag, 1986, $98 set [ISBN: 0-387-96257-3]. Part 1: Introduction, Text, and Translation, x + 375 pp; Part 2: Commentary, Index, and Figures, vi + 371 pp. Another scholarly production in the publisher's Sources in the History of Mathematics and Physical Sciences Series. Book 7 of Pappus' Collection has been an important source of information concerning lost works of Euclid and Apollonius and has served as an inspiration for later workers in geometry. This edition is a revision of the author's doctoral dissertation. JK

History, S, P, L. R.A. Fisher: The Life of a Scientist. Joan Fisher Box. Wiley, 1978, xii + 512 pp, $22.50 (P). [ISBN: 0-471-09300-9] Paperback reprint of the 1978 original hardcover edition (TR, April 1979). LAS

Foundations, P. Space, Time, and Life: The Probabilistic Pathways of Evolution. V.V. Nalimov. Transl: A.V. Yarkho. ISI Pr, 1985, xix + 110 pp, $24.95. [ISBN: 0-89495-048-7] A modern metaphysics of numbers, translated from Russian, relying on diverse quotations from philosophers, scientists, and theologians and numerous plays on words to suggest a new numerology in which "number" is the organizing principle of the world. Fourth in a series of philosophical works by Nalimov, all published by ISI Press. LAS

Combinatorics, P. Structure of the Standard Modules for the Affine Lie Algebra $A_1^{(1)}$. James Leopwsky, Mirko Primc. Contemp. Math., V. 46. AMS, 1985, ix + 84 pp, $13 (P). [ISBN: 0-8218-5048-2] $A_1^{(1)}$ is the Kac-Moody Lie algebra for sl(2), the three-dimensional complex Lie algebra of 2x2 trace-zero matrices. Its representation theory serves as the prototype for other infinite-dimensional Lie algebras. Analysis of the structure leads to combinatorial "Rogers-Ramanujan" identities. BC

Discrete Mathematics, T*(13: 1), S, L. Discrete Structures: An Introduction to Mathematics for Computer Science. Fletcher R. Norris. Prentice-Hall, 1985, xiv + 321 pp, $32.95. [ISBN: 0-13-215260-6] A first course in discrete mathematics which emphasizes applications to computer science. Includes introductory material in logic, sets, Boolean algebra, circuits, recursion, induction, graphs, counting and posets. Readable with an adequate supply of exercises. CEC

Number Theory, S(18), P. Number Theory and Combinatorics, Japan 1984. Ed: Jin Akiyama, et al. World Scientific, 1985, viii + 446 pp, $56. [ISBN: 9971-978-77-6] A collection of twenty-six papers dealing with analytic number theory and with connections between number theory, analysis, and combinatorics. These make up the proceedings of three conferences which coincided with Paul Erdös' visit to Japan in January of 1984. CEC

Number Theory, S(16-18), P. Prime Numbers and Computer Methods for Factorization. Hans Riesel. Prog. in Math., V. 57. Birkhauser Boston, 1985, xvi + 464 pp, $44.95. [ISBN: 0-8176-3291-3] Factoring numbers is easy, but doing it quickly is another question. This book collects numerous algorithms and results on the distribution and recognition of primes and the factorization of everything else. Also contains PASCAL code and instructions for doing hundred-digit arithmetic and fast multiplication. BC

Number Theory, T*(14: 1, 2), S, P, L*. Solved and Unsolved Problems in Number Theory, Third Edition. Daniel Shanks. Chelsea, 1985, xiv + 304 pp, $18.95. [ISBN: 0-8284-1297-9] Of the four chapters in this volume, the first three are essentially identical to the previous volume (TR, June-July 1979). Chapter four which is entitled "Progress" has almost been doubled in length because of new developments. CEC

Number Theory, S(17), P. Topics in Analytic Number Theory. Ed: Sidney W. Graham, Jeffrey D. Vaaler. U of Texas Pr, 1985, vii + 312 pp, $35. [ISBN: 0-292-75530-9] The proceedings of a conference held in the summer of 1982 at the University of Texas in Austin. Includes long papers by Iwaniec and Halberstrom and shorter papers by eleven other participants. Topics cover a broad range of current research. CEC

Number Theory, T(16: 1), S, P, L. Prime Numbers. William and Fern Ellison. Wiley, 1985, xii + 417 pp, $49.95 (P). [ISBN: 0-471-82653-7] A translation of the 1975 text Les Nombres Premiers (TR, March 1976). Still an excellent introduction to the theory of prime numbers with a superb collection of exercises. Unfortunately, it has not been updated. CEC

Linear Algebra, T(13-14: 1). Introduction to Linear Algebra, Second Edition. Serge Lang. Undergrad. Texts in Math. Springer-Verlag, 1986, viii + 291 pp, $34. [ISBN: 0-387-96205-0] Second Edition of a generally conventional textbook. Differs from the First Edition (1970) chiefly in having a chapter on eigenvectors and eigenvalues and a less complete treatment of determinants. JD-B

Linear Algebra, T(14-15: 1), L. Linear Algebra with Applications, Second Edition. Steven J. Leon. Macmillan, 1986, xv + 408 pp. [ISBN: 0-02-369810-1] Mostly an expansion of the First Edition (TR, February 1982); most notable additions are introduction of the exponential matrix and more material on positive definite matrices. JS

Linear Algebra, T(17: 1). Multilinear Algebra. D.G. Northcott. Cambridge U Pr, 1984, x + 198 pp, $39.50. [ISBN: 0-521-26269-0] Graduate-level text covering multilinear mappings; tensor, symmetric, exterior, and Hopf algebras; and graded duality. Assumes knowledge of modules. Exercises at end of each chapter with complete solutions for most important ones. KS

Linear Algebra, T(13-14: 1).** Introduction to Linear Algebra. Robert F.V. Anderson. Holt, Rinehart & Winston, 1986, xix + 388 pp, $31.95. [ISBN: 0-03-921835-X] Attractive treatment of the usual material. Clearly written. Profusely exampled. Plenty of straightforward exercises plus a good supply of non-routine problems which extend the text. Lots of motivational tips. Flexible enough for a variety of course needs. Chapters on numerical methods and linear programming. Brief but rather nice chapter on matrices and geometry. JK

Linear Algebra, C(13-14). Linear Algebra Computer Companion. Apple II. Gareth and Donna Williams. Allyn & Bacon, 1984, $35. An integrated package of program disk, data disk, and student manual. Designed to accompany Linear Algebra with Applications (Allyn & Bacon, 1984), but functions nicely as stand-alone toolkit for student and teacher. Includes programs for matrix multiplication, powers, transformations, inverses, solving linear systems, normalized vectors, orthogonalization, eigenvalues and eigenvectors, simplex method, Markov chains, digraphs, and a 4-program graphics package. User options allow for screen scrolling (40-column display limits visibility of large matrices), data disk input, and output to printer, if desired. Manual provides explanations, examples, and problems, as well as keystroke sequences and controls. Menu-driven programs work on one or two drive systems. RD

Group Theory, P. Finite Groups--Coming of Age. Ed: John McKay. Contemp. Math., V. 45. AMS, 1985, x + 350 pp, $30 (P). [ISBN: 0-8218-5047-4] Proceedings of the Canadian Mathematical Society Conference held at Concordia University, Montreal, June 15-28, 1982, concerning finite group theory which has not been central to the classification of finite simple groups. Seventeen papers. DFA

Group Theory, P. Finite Groups of Lie Type: Conjugacy Classes and Complex Characters. Roger W. Carter. Wiley, 1985, xii + 544 pp, $69.95. [ISBN: 0-471-90554-2] Self-contained development (modulo results on reductive algebraic groups and l-adic cohomology, which are summarized) of the Deligne-Lusztig theory of irreducible characters for finite groups of Lie type. Suitable as an advanced introduction; most valuable as a reference. BC

Calculus, T(15-16: 1). Analysis I. Wolfgang Walter. Grund. Math., B. 3. Springer-Verlag, 1985, xii + 385 pp, DM 48 (P). [ISBN: 0-387-12780-1] A leisurely but careful text (in German) on the calculus of functions of one real variable. Many historical remarks, exercises. JD-B

Calculus, C(12-14). Compucalc: The Computer Calculus. Apple II. Robert J. Weaver. Worth, 1984, $18.75 (disk), $8.95 (workbook). Program disk and workbook designed for use with Calculus (Munem and Foulis, Worth), but can be used with any standard calculus text. Designed for numerical (not symbolic) manipulation, its purpose is to give student and teacher a high-speed calculator and graphics toolkit. Provides function evaluation, search for limits, derivative and integral calculations, and chain rule, Cartesian, polar, and parametric graphing. Not intended for programmed learning. Operates quickly and clearly with straightforward commands. Function input follows syntax of Basic, with menu-driven programs that allow user to retain a specific function from program to program, a time-saving feature. Requires 64K memory and one disk drive; printer optional. IBM version also available. RD

Calculus, T(13: 1), S, L. Calculus and the Personal Computer. C.H. Edwards, Jr. Prentice-Hall, 1986, viii + 200 pp, $13.95 (P). [ISBN: 0-13-112319-X] A collection of IBM-PC BASIC programs and associated exercises that introduce and explore numerical methods of elementary calculus, e.g., numerical integration, numerical solution of equations, Euler's method. Graphics capability is exploited. Error estimation is not emphasized. Applications are chiefly to physical modelling. PZ

Calculus, S(13-14). Higher Mathematics for Engineering Students, Part 1: Linear Algebra and Fundamentals of Mathematical Analysis. Ed: A.V. Efimov, B.P. Demidovich. Transl: Leonid Levant. MIR (US Distr: Imported Pub), 1984, 511 pp, $9.95. [ISBN: 0-8285-2890-X] Roughly comparable to the Schaum Outline Series in purpose and format, this first of three volumes deals with problems using topics of the first-year mathematics for Russian engineering students. Includes basic linear algebra, determinants, differential and integral one-variable calculus, some partial derivatives. Some textual explanation, examples, hints, and answers. JS

Real Analysis, T*(15: 1). Introduction to Convergence. S.C. Malik. Wiley, 1984, vii + 210 pp, $26.95. [ISBN: 0-470-20070-7] Sequences, series and products covered in depth (e.g., Raabe's and Kummer's tests for series). Classical in style with a large collection of examples, lesser number of exercises. Well written. TAV

Complex Analysis, P. Selecta, Donald C. Spencer. Donald C. Spencer. World Scientific, 1985, $179. Volume 1, xii + 656 pp; Volume 2, viii + 654 pp; Volume 3, viii + 437 pp. Most of Spencer's 90 papers, organized into ten subject areas--mostly in conformal mappings, deformation of structures on manifolds, differential operators and Lie equations. Includes a brief biography by J.J. Kohn, a complete bibliography of Spencer's work, and a list of Ph.D. theses that he supervised. LAS

Complex Analysis, P. Bibliography of Schlicht Functions: Part I (-1965), Part II (1966-1975), Part III (1976-1981). S.D. Bernardi. Mariner, 1982, x + 353 pp. [ISBN: 0-936166-09-6] Lists over 2500 references (books, journals, technical reports, etc.) concerning theory of analytic univalent functions. In three parts: up to 1965, 1966-75, 1976-81. Items in Part I and II are further classified by subject. The main theorem in this area--the Bieberbach conjecture--was proved in 1984. (1966 New York Univ. Ed., Part I, TR, August-September 1967; 1977, Part II, TR, January 1978.) PZ

Complex Analysis, P*. Papers on Fuchsian Functions. Henri Poincaré. Transl: J. Stillwell. Springer-Verlag, 1985, 483 pp, $34. [ISBN: 0-387-96215-8] Five major papers on discontinuous groups and automorphic functions. Written over 100 years ago, they remain valuable reading, and the translation is welcome. Springer-Verlag should be ashamed, though, of the poor-quality (computer produced?) typeface; Poincaré deserves better. BC

Differential Equations, T(14: 1). Ordinary Differential Equations with Linear Algebra. David Lomen, James Mark. Prentice-Hall, 1986, xi + 350 pp, $29.95. [ISBN: 0-13-639782-4-01] Presupposes only calculus. The first, preliminary chapter (over 20% of the book) treats elementary linear algebra. A last, brief chapter mentions numerical methods. In between are the usual first-course topics. Short sections, many drill problems. Attention to applications. DFA

Differential Equations, S(15-17), P, L*. Chaos. Hao Bai-Lin. World Scientific, 1984, 576 pp, $26 (P); $56. [ISBN: 9971-966-51-4; 9971-966-50-6] A very useful volume: a 75-page exposition of the occurrence of turbulence and chaos, especially in physical systems, followed by reprints of 41 key papers providing historical, mathematical, and (primarily) physical background. Concludes with a comprehensive bibliography. LAS

Differential Equations, T*(14), L. Ordinary Differential Equations with Numerical Techniques. John L. Van Iwaarden. Harcourt Brace Jovanovich, 1985, ix + 530 pp, $27.95. [ISBN: 0-15-567550-8] Very attractive and readable. Large selection of examples and problems including many which concern real-world applications. Half of the ten chapters are devoted to numerical techniques. A very nice treatment of the standard first-course topics which integrates analytical and numerical methods. DFA

Differential Equations, T(14: 2). Differential Equations with Linear Algebra. Zbigniew H. Nitecki, Martin M. Guterman. Saunders College, 1986, xi + 644 pp, $34.95. [ISBN: 0-03-002719-5] For a course combining the usual first course in ordinary differential equations and that in linear algebra. Presupposes only the calculus sequence. Uses physical models to motivate the study of differential equations; uses differential equations to motivate the study of linear algebra. A multitude of examples and exercises. For students of engineering, science, and mathematics. DFA

Differential Equations, T(16-17: 1, 2), S, L. Singular-perturbation Theory: An Introduction with Applications. Donald R. Smith. Cambridge U Pr, 1985, xvi + 500 pp, $42.50. [ISBN: 0-521-30042-8] Suitable for students of mathematics, pure and applied sciences, engineering. Studies initial-value problems of oscillatory type (using methods of averaging and multiple scales), of overdamped type (O'Malley-Hoppensteadt and matched expansions methods), and boundary value problems (multivariable method). Numerous applications from areas of science and engineering; many references to the literature. DFA

Differential Equations, T*(14: 1). Introduction to Differential Equations: ODE, PDE, and Series. Richard E. Williamson. Prentice-Hall, 1986, xiii + 443 pp, $31.95. [ISBN: 0-13-480989-0] Assumes only calculus. Standard first-course topics, including first-order systems, a good treatment of linear systemss, the prerequisite linear algebra, plus a chapter on linear partial differential equations. Applications and numerical techniques throughout. Encourages geometric viewpoint. An attractive book which deserves consideration. DFA

Partial Differential Equations, S(16-18), P. Lecture Notes in Mathematics-1150: Rearrangements and Convexity of Level Sets in PDE. Bernhard Kawohl. Springer-Verlag, 1985, 136 pp, $9.80 (P). [ISBN: 0-387-15693-3] A rearrangement is a type of transformation of the admissible functions of a variational problem. Rearrangements which leave a variational functional invariant may be used to establish regularity and symmetry properties of solutions to variational problems. This book discusses applications of rearrangements to problems in semilinear differential equations. AM

Partial Differential Equations, T(18: 1), S, P. Multiple Time Scales. Ed: Jeremiah U. Brackbill, Bruce I. Cohen. Comput. Tech., V. 3. Academic Pr, 1985, xiii + 442 pp, $75. [ISBN: 0-12-123420-7] This volume consists of thirteen chapters, each one written by a different author or group of authors. The chapters include an introductory overview along with a variety of applications of using large computers to solve problems possessing wide ranges of time scales. Successful numerical methods for each application are described. CEC

Partial Differential Equations, P. Pseudodifferential Operators and Applications. Ed: Francois Trèves. Proc. of Symp. in Pure Math., V. 43. AMS, 1985, vii + 300 pp, $44. [ISBN: 0-8218-1469-9] Nineteen papers devoted to microlocal analysis, the precise application of Fourier transforms to analysis on manifolds. From an AMS symposium held at the University of Notre Dame, April 2-5, 1984. DFA

Partial Differential Equations, P. Hypoellipticité Maximale pour des Opérateurs Polynomes de Champs de Vecteurs. Bernard Helffer, Jean Nourrigat. Prog. in Math., V. 58. Birkhauser Boston, 1985, ix + 278 pp, $34.95. [ISBN: 0-8176-3310-3] Concerns regularity theory of partial differential operators constructed from certain systems of vector fields. Includes new results, expanded proofs of previously published results, and a central conjecture in the field. Applications include problems in several complex variables. PZ

Numerical Analysis, S(16-18), P. Finite Element Methods in Linear Ideal Magnetohydrodynamics. Ralf Gruber, Jacques Rappaz. Ser. in Comput. Physics. Springer-Verlag, 1985, xi + 180 pp, $34. [ISBN: 0-387-13398-4] The finite element method is a numerical technique for approximating the solutions

and eigenvalues for a Sturm-Liouville problem. It is based on the fact that the eigenvalues of a self adjoint linear operator are the critical values of its Raleigh-quotient. The book discusses numerical problems that arise when the finite element method is used to study instabilities of thermonuclear plasmas. AM

Numerical Analysis, P. Stiff Computation. Ed: Richard C. Aiken. Oxford U Pr, 1985, xiv + 462 pp, $75. [ISBN: 0-19-503453-8] Contributions stemming from a conference at Park City, Utah, in 1982 to review work on "stiff" computation. Includes a discussion of computer packages for solving stiff ordinary differential equations. (Stiffness is an as-yet imprecise quality implying severe restrictions on standard numerical techniques.) BC

Numerical Analysis, P. The Numerical Solution of Nonlinear Stiff Initial Value Problems: An Analysis of One Step Methods. W.H. Hundsdorfer. CWI Tract, V. 12. Math Centrum, 1985, 138 pp, Dfl. 20,30 (P). [ISBN: 90-6196-283-8] Reprint of the author's thesis written at the University of Leiden under M.N. Spijker. Runge-Kutta methods and generalizations; existence of unique solutions to the algebraic equations in implicit and semi-implicit methods; contractivity and error propagation per step; B-convergence for several θ-methods. DFA

Numerical Analysis, P. Boundary Elements VII. Ed: C.A. Brebbia, G. Maier. Springer-Verlag, 1985, $159.50 set [ISBN: 0-387-15729-8]. Volume I, 664 pp; Volume II, 634 pp. Proceedings of a September 1985 international conference at Lake Como, Italy--the seventh in a series. Papers address issues of accuracy, reliability, and diagnostic tools in the various computer codes that implement boundary element methods, primarily in engineering problems. LAS

Functional Analysis, P. Lecture Notes in Mathematics-1132: Operator Algebras and their Connections with Topology and Ergodic Theory. Ed: H. Araki. Springer-Verlag, 1985, vi + 594 pp, $35.50 (P). [ISBN: 0-387-15643-7] Proceedings of a conference held in Busteni, Romania, August 29-September 9, 1983. JAS

Functional Analysis, P. Lecture Notes in Mathematics-1166: Banach Spaces. Ed: N. Kalton, E. Saab. Springer-Verlag, 1985, vi + 199 pp, $14.40 (P). [ISBN: 0-387-16051-5] 22 contributed papers from the June 1984 CBMS Regional Conference held at the University of Missouri, intended to complement the invited lectures by Gilles Pisier which appeared in the CBMS Regional Conference Series published by AMS. LAS

Functional Analysis, S(17-18), P. Generators of Strongly Continuous Semigroups. J.A. van Casteren. Res. Notes in Math., V. 115. Pitman, 1985, 203 pp, $17.95 (P). [ISBN: 0-273-08669-3] If X is a Banach space, a semigroup on X is a one parameter family of continuous linear maps $P(t)$, $t \geq 0$ satisfying $P(0) = I$; $P(s+t) = P(s)\,P(t)$; and $\lim_{t\to 0} P(t)x = x$, $x \in X$. Book describes recent results in the theory of semigroups as well as applications to Schrodinger operators, positivity preserving semigroups, quadratic form theory, and Feynman path integrals. Contains descriptions of several open problems. AM

Functional Analysis, P. Theory of Bases and Cones. P.K. Kamthan, M. Gupta. Research Notes in Math., V. 117. Pitman, 1985, 256 pp, $17.95 (P). [ISBN: 0-273-08657-X] Intended for readers familiar with locally convex spaces, this book begins with a brief review (no proofs), then in greater detail covers much of Schauder basis theory and the geometric interpretation of the theory. Densely written with few exercises and little discussion of the many theorems proved. MU

Functional Analysis, P. Factorization of Linear Operators and Geometry of Banach Spaces. Gilles Pisier. CBMS Reg. Conf. Ser. in Math., No. 60. AMS, 1986, x + 154 pp, $18 (P). [ISBN: 0-8218-0710-2] Lectures delivered at the June 1984 CBMS Regional Conference held at the University of Missouri on contributions since 1968 to the general question of when an operator between two Banach spaces factors through a Hilbert space. Major focus is on extensions and applications of Grothendieck's 1956 theorem on the metric theory of tensor products. Contributed papers are published separately (Springer LNM-1166). LAS

Functional Analysis, P. Lecture Notes in Mathematics-1128: Singular Ordinary Differential Operators and Pseudodifferential Equations. Johannes Elschner. Springer-Verlag, 1985, 200 pp, $14.40 (P). [ISBN: 0-387-15194-X] Fuchsian differential operators with one singular point; linear ordinary differential operators with one singular point and with several; elliptic differential operators in R^n degenerating at one point; degenerate pseudodifferential operators on a closed curve; and a finite element method for pseudodifferential equations on a closed curve. DFA

Functional Analysis, T(18: 2), S, P. Nonlinear Functional Analysis and its Applications, I: Fixed-Point Theorems. Eberhard Zeidler. Transl: Peter R. Wadsack. Springer-Verlag, 1986, xxi + 897 pp, $98. [ISBN: 0-387-90914-1] Part I of a massive and comprehensive study of nonlinear functional analysis and applications. This volume consists of three parts: fundamental fixed-point principles, applications, and the mapping degree and fixed-point index. Problems and references for each chapter. Various summary lists, index, extensive bibliography. JS

Functional Analysis, P. Lecture Notes in Mathematics-1153: Probability in Banach Spaces V. Ed: A. Beck, et al. Springer-Verlag, 1985, vi + 457 pp, $29.20 (P). [ISBN: 0-387-15704-2] Proceedings of the fifth in a biennial series of conferences that focuses on issues of common interest to probabilists, statisticians, functional analysts and harmonic analysts. Contains 25 research reports. LAS

Functional Analysis, S(18), P. Sobolev Spaces. Vladimir G. Maz'ja. Transl: T.O. Saposnikova. Springer-Verlag, 1985, xix + 486 pp, $59. [ISBN: 0-387-13589-8] A fairly self-contained discussion of Sobolev spaces starting from basic properties and emphasizing imbedding theorems. Applications to boundary value problems for elliptic operators. Comments conclude each chapter. Indexes, extensive bibliography. JS

Analysis, T(17: 1), S. Integration. Hans Günzler. Bibliographisches Inst, 1985, 392 pp, (P). [ISBN: 3-411-03101-8] A brief, sophisticated textbook (in German) on the abstract Lebesgue integral. Many exercises, half with hints on solutions. JD-B

Analysis, S(18), P. Analysis in Classes of Discontinuous Functions and Equations of Mathematical Physics. A.I. Vol'pert, S.I. Hudjaev. Mechanics: Analysis, V. 8. Martinus Nijhoff, 1985, xviii + 678 pp, $117.50. [ISBN: 90-247-3109-7] Treats theory and applications of analysis in BV-spaces, i.e., spaces of functions whose distribution derivatives are measures. Such functions, though not necessarily continuous, are regular enough to allow derivative operations. Problems in differential and integral equations, mathematical physics and chemical physics are studied in the BV-space context. A long first chapter develops the functional analysis and measure theory used elsewhere. A lot of mathematics in one package, but note the price. PZ

Analysis, S(14-15). Higher Mathematics for Engineering Students, Part 2: Advanced Topics of Mathematical Analysis. Ed: A.V. Efimov, B.P. Demidovich. Transl: Leonid Levant. MIR (US Distr: Imported Pub), 1984, 414 pp, $9.95. [ISBN: 0-8285-2891-8] A continuation intended for second-year Russian engineering students. Problems deal with multiple integration, differential equations, vector analysis, complex variables, operational calculus. Some text, examples, hints, answers. JS

Analysis, T*(16-17: 2), P*, L*. Constructive Analysis. Errett Bishop, Douglas Bridges. Grund. der math. Wissenschaften, B. 279. Springer-Verlag, 1985, xii + 477 pp, $48. [ISBN: 0-387-15066-8] Substantially based on Bishop's 1967 work Foundations of Constructive Analysis (TR, May 1968; Extended Review, October 1968), this book contains much new material. Begins with "A Constructivist Manifesto" where the ground rules are established. An important contribution to the fields of real and functional analysis. TAV

Differential Geometry, P. Lecture Notes in Mathematics-1165: Seminar on Deformations. Ed: J. Ławrynowicz. Springer-Verlag, 1985, ix + 331 pp, $20.50 (P). [ISBN: 0-387-16050-7] 22 research papers from the 1982-84 seminar organized in Łódź, and conducted in part at four spontaneous seminars at the 1983 International Congress in Warsaw. LAS

Differential Geometry, P. Differential Geometry, Calculus of Variations, and Their Applications. Ed: George M. Rassias, Themistocles M. Rassias. Lect. Notes in Pure & Appl. Math., V. 100. Dekker, 1985, xiii + 521 pp, $75 (P). [ISBN: 0-8247-7267-9] Collection of 33 papers in differential geometry concentrating on calculus of variations. Topics addressed include Lagrangian formalisms, Yang-Mills theory, gauge field theories, conservation laws, Morse-Smale index theorem for minimal surfaces, Padé approximants. AM

Differential Geometry, P. Invariance Theory, The Heat Equation, and the Atiyah-Singer Index Theorem. Peter B. Gilkey. Publish or Perish, 1984, viii + 349 pp, $40. [ISBN: 0-914098-20-9] Develops and proves the Atiyah-Singer index theorem in complete generality, using heat equation methods. Necessary differential analysis for defining and computing the index of an elliptic operator is reviewed. Invariance theory and differential geometry--also reviewed in the text--link the analytic and topological viewpoints. PZ

Differential Geometry, P. Lecture Notes in Mathematics-1156: Global Differential Geometry and Global Analysis 1984. Ed: D. Ferus, et al. Springer-Verlag, 1985, 339 pp, $20.50 (P). [ISBN: 0-387-15994-0] Proceedings of a June 1984 international colloquium in Berlin; 21 research contributions. LAS

Algebraic Topology, P. Lecture Notes in Mathematics-1126: Algebraic and Geometric Topology. Ed: A. Ranicki, N. Levitt, F. Quinn. Springer-Verlag, 1985, 423 pp, $25.80 (P). [ISBN: 0-387-15235-0] Proceedings of a conference held at Rutgers University, New Brunswick, USA, July 6-13, 1983. JAS

Topology, T(18: 2), P. Topological Vector Spaces. Lawrence Narici, Edward Beckenstein. Pure & Appl. Math., V. 95. Dekker, 1985, xii + 408 pp, $69.75. [ISBN: 0-8247-7315-2] After a development of filters, commutative topological groups and completeness, the authors develop TVS-theory including duality, barreled, and bornological spaces, reflexivity, and approximation. Clearly written and very complete. TAV

Optimization, T(15). Introduction to Linear and Convex Programming. Neil Cameron. Australian Math. Soc. Lect. Ser., V. 1. Cambridge U Pr, 1985, 149 pp, $34.50; $12.95 (P). [ISBN: 0-521-30951-4; 0-521-31207-8] A brief introduction to linear programming using the simplex algorithm, and aspects of convex programming including the use of Fenchel transforms. Suitable for an overview of the ideas. TAV

Optimization, P, L. Nonlinear Optimization Bibliography with two-level Key-word and Author Indexes. Gerald Berman. U of Waterloo Pr, 1985, 312 pp, $50 (P). [ISBN: 0-88898-046-9] 4000 articles from journals and conference proceedings, primarily from Mathematical Reviews, up to 1984. Section 1 contains a complete bibliographic index with an eight-character mnemonic code that uniquely identifies each article. Section 2 is a two-level listing of keywords from the titles together with

corresponding mnemonic codes. Section 3 is an author index to permit one to find papers with multiple authors. LAS

Dynamical Systems, P. Averaging Methods in Nonlinear Dynamical Systems. J.A. Sanders, F. Verhulst. Appl. Math. Sci., V. 59. Springer-Verlag, 1985, x + 247 pp, $28 (P). [ISBN: 0-387-96229-8] A thorough, introductory treatment of asymptotic analysis of nonlinear dynamical systems, emphasizing the method of averaging. Concludes with normal forms for Hamiltonian systems. Appendices cover historical origins and necessary background from global analysis. LAS

Dynamical Systems, P. Lecture Notes in Mathematics-1125: Dynamical Systems and Bifurcations. Ed: B.L.J. Braaksma, H.W. Broer, F. Takens. Springer-Verlag, 1985, 129 pp, $9.80 (P). [ISBN: 0-387-15233-4] Proceedings of a workshop held in Groningen, The Netherlands, April 16-20, 1984. JAS

Control Theory, P. Lecture Notes in Mathematics-1154: Singular Perturbation Analysis of Discrete Control Systems. D.S. Naidu, A.K. Rao. Springer-Verlag, 1985, x + 195 pp, $14.40 (P). [ISBN: 0-387-15981-9] Singular perturbation of difference equations in classical and state-space form; applications to open- and closed-loop optimal control. Many explicit examples, tables, and graphs. BC

Control Theory, T(16-17: 1). Control Systems Modeling and Analysis. Gerard Voland. Prentice-Hall, 1986, xiv + 266 pp, $37.95. [ISBN: 0-13-171984-X] Introduction for systems engineers to control theory, emphasizing differential equations and the Laplace transform. Many exercises and solved examples. Solution-oriented, with little theory and no proofs. BC

Control Theory, P. Control and Dynamic Systems: Advances in Theory and Applications, Volume 22: Decentralized/Distributed Control and Dynamic Systems, Part 1 of 3. Ed: C.T. Leondes. Academic Pr, 1985, xx + 398 pp, $55. [ISBN: 0-12-012722-9] A collection of ten papers with two more volumes to follow. JAS

Systems Theory, P. Lecture Notes in Economics and Mathematical Systems-257: Dynamics of Macrosystems. Ed: J.-P. Aubin, D. Saari, K. Sigmund. Springer-Verlag, 1985, vi + 279 pp, $20.50 (P). [ISBN: 0-387-15987-8] Proceedings of a September 1984 conference held in Luxemburg, Austria covering a wide range of dynamical systems from biological to socioeconomic. LAS

Statistics, P. Teaching of Statistics in the Computer Age. Ed: Lennart Råde, Terry Speed. Chartwell-Bratt, 1985, 244 pp. [ISBN: 0-86238-090-1] Proceedings from Round Table Meeting on teaching statistics, associated with Fifth International Congress on Mathematics Education, Canberra, Australia, August 1984. Includes conference recommendations and seventeen papers. Focus is on integration of calculators and computers into the teaching of statistics; includes curriculum issues, course and program descriptions, and requisite skills for new age of learning statistics. Extensive bibliography. RD

Statistics, C(14-16). StatView. Macintosh. Dan Feldman, Jim Gagnon. Brainpower (24009 Ventura Blvd., Suite 250, Calabasas, CA 91302 (818)884-6911), 1985, $189.95. A statistics package for those with good knowledge of statistics, consisting of program disk and reference manual (112 pages). Manual would benefit from addition of an index. Package includes descriptive statistics (arithmetic, geometric and harmonic mean; standard deviation; median; mode; kurtosis and skewness; frequency distributions), comparative statistics (t-test; correlation; simple, polynomial and multiple regression; analysis of variance; chi-square tests; nonparametric statistics; etc.), and graphics (scatter-plots, bar charts, pie charts and line charts). Data are entered as columns: the number of rows and columns permitted is determined by size of memory. Has some spreadsheet-type capabilities. Files may be imported from other sources. Not copy protected. Runs on 128K, but better on 512K. KK

Computer Literacy, S(13-14). Computer Work Stations. Herman Holtz. Chapman & Hall, 1985, xv + 302 pp, $24.50. [ISBN: 0-412-00491-7] Easy, clear reading with lots of elementary information for an introduction. The scope is limited and dated, however, so don't depend on this for up-to-date details or sufficient information to avoid serious problems, e.g., details of RS-232C standard are not covered, new (1982) fast storage cartridge systems are not covered. The emphasis is on popular microcomputers, ignoring Sun, Apollo or Convergent Technologies level. Nothing on laser printers, 2400 baud (and up) modems. Read this first, not last. It does introduce the basic ideas, suggests check lists, and covers an appropriately wide variety of issues. No mention of languages newer than the middle 1970's. JAS

Computer Literacy. Computing Information Directory: A Comprehensive Guide to the Computing Literature. Ed: Darlene Myers Hildebrandt. Pedaro Inc (32606 7th Ave. SW, Federal Way, WA 98023), 1985, v + 557 pp, $99.95 (P). [ISBN: 0-933113-00-5] A listing of various information sources which is difficult to describe because of the inherent disorganization of its topic. Various chapters provide lists of journals (all kinds), university computer center newsletters, languages, books, technical reports as well as software reviews and hardware sources. Reasonably comprehensive, mildly eclectic (neither C nor LISP occurs in the short list of computer languages), but probably a helpful place to start looking. This 1985 edition is a successor to the 1981 Computer Sciences Resources. JAS

Computer Literacy, S. The Easy Guide to Your Macintosh. Joseph Caggiano. Sybex, 1984, xviii + 214 pp, (P). [ISBN: 0-89588-216-7] A very general guide to using the Macintosh. Contains much less than the documents that come with a Macintosh. LAS

Computer Literacy, P, L. The Human Factor: Designing Computer Systems for People. Richard Rubinstein, Harry M. Hersh. DEC, 1984, 249 pp, $25 (P). [ISBN: 0-932376-44-4] About 100 guidelines for designing "user-friendly" software. Many good examples merely illustrate common sense. LAS

Computer Literacy, S(13-16). Computer: A Challenge for Business Administration. A.-W. Scheer. Springer-Verlag, 1985, x + 256 pp, $19.50. [ISBN: 0-387-15514-7] A survey, halfway between computer literacy and business administration. Certainly better done than most of the genre, but not heavy on conceptual material. Not designed as a text. JAS

Computer Literacy, S(13). The Incredible Shrinking Computer: A Guide to Computer Literacy. John G. Sancin, Jr. DEC, 1984, 166 pp, $21 (P). [ISBN: 0-932376-30-4] Elementary guide to personal computers. LAS

Computer Literacy, S(13-15), L. Cohabiting with Computers. Ed: Joseph F. Traub. William Kaufmann, 1985, x + 185 pp, $15. [ISBN: 0-86576-079-9] Speculations by outstanding scientists (including Lewis Branscomb, Edward David, Arno Penzias, Herbert Simon) on the future role of computers in society, based on addresses given at Columbia University in October 1983 at the dedication of their new computer science building. LAS

Computer Programming, S. The Macintosh Basic Handbook. Thomas Blackadar, Jonathan Kamin. Sybex, 1984, $24.95 (P). [ISBN: 0-89588-257-4] A comprehensive alphabetic compendium of all 189 keywords in Macintosh Basic together with the most commonly used of the 300 or so Toolbox commands. Each entry contains syntax details, examples of use (often with complete illustrative programs) and applications. LAS

Computer Programming, T?(13-14). Structured COBOL Programming. Robert T. Grauer. Prentice-Hall, 1985, xvi + 479 pp, $22.95 (P). [ISBN: 0-13-854217-1-01] A conventional programming text covering both COBOL 74 and COBOL 8X. Punch cards are abandoned for text editions, but flowcharts co-exist with pseudocode. JAS

Computer Programming, S(13-14), P, L. Create Your Own Games Computers Play. Keith S. Reid-Green. DEC, 1984, x + 242 pp, $21 (P). [ISBN: 0-932376-29-0] A serious introduction in structured Basic to sophisticated computer graphics: animation, randomness, mazes, computer-aided design. Contains good, practical introduction to perspective, rotations, and other algorithms necessary to produce realistic high-quality game graphics. LAS

Computer Programming, T(1), S. C for Personal Computers: IBM PC, AT&T PC 6300, and Compatibles. Narain Gehani. Principles of Comp. Sci. Ser. Computer Science Pr, 1985, xiii + 301 pp, $19.95 (P). [ISBN: 0-88175-111-1] In spite of the title, this is a generic C book; the few PC specific examples could easily be adapted to an arbitrary computer. Written for readers with a good knowledge of at least one procedural programming language, the book is fast-paced and provides interesting nontrivial examples. Except for some minor non-UNIX idiosyncrisies in the library functions presented with lattice/Microsoft C, the book treats C as a global portable language discussing the emergent ANSI standard. JAS

Computer Programming, T(13-14: 1), S. Macintosh Pascal. Lowell A. Carmony, Robert L. Holliday. Comp. & Math. Ser. Computer Science Pr, 1985, xiii + 497 pp, $19.95 (P). [ISBN: 0-88175-081-5] An introduction to Pascal on the Macintosh, written and typeset on a Mac, with optional disc containing all example programs. Includes nonstandard Pascal features (e.g., string types, graphics) in the core text, but with warnings. Discusses records, files, recursion, and linked lists. LAS

Computer Programming, T(13: 1), S. Learning Apple FORTRAN. Donald J. Geenen. Comp. & Math. Ser. Computer Science Pr, 1986, xiii + 264 pp, $17.95 (P). [ISBN: 0-88175-024-7] For high school and beginning college students, written by a high school teacher. Suitable for self-study, the first third of the book is on the Apple FORTRAN operating system, the rest on the Apple version of FORTRAN 77. Includes chapters on the Applestuff and Turtlegraphics library units and word processing. Review questions and answers; programming problems. DFA

Computer Programming, T. Commodore 64 Assembly Language: A Course of Study Based on the DEVELOP-64 Assembler/Editor/Debugger. W. Douglas Maurer. Computer Science Pr, 1985, xiv + 415 pp, $19.95 (P). [ISBN: 0-88175-040-9] A rather machine- and assembler-specific book. Presents several worthwhile algorithms but often not in a larger context. It might be hard for a student to generalize from this approach, which is better directed at the hobbiest than the computer science student. Presumes previous experience with BASIC or Pascal. JAS

Computer Programming, T(13-14: 1), S. Hands-on Basic for the DEC Professional. Herbert Peckham, Wade Ellis, Jr., Ed Lodi. DEC, 1985, xiii + 333 pp, $22 (P). [ISBN: 0-932376-66-5] A revision of the popular 1977 text Hands-on-Basic adapted specifically to the DEC Professional desk-top computer. Covers strings, files, and graphics, as well as the usual topics. Many exercises, with some answers (often full programs). DEC's lawyers have no confidence, however, proclaiming in the Foreward that "we cannot guarantee the correctness or accuracy of its contents." LAS

Computer Programming, T(13: 1). Learning Pascal Step by Step. Vern McDermott, Andrew Young, Diana Fisher. Computer Science Pr, 1985, xv + 236 pp, $19.95. [ISBN: 0-88175-046-8] Written in 24 lessons, each containing these sections: objectives, introduction of concept, demonstration problems, skill development exercises. For class or self-study. Suitable for high school and junior college students. Machine specific for Apple USCD Pascal, TRS-80 New Classics Pascal-80, IBM P-System Ver-

sion IV.0. DFA

Computer Programming, S. Using the Macintosh Toolbox with C. Fred A. Huxham, David Burnard, Jim Takatsuka. Sybex, 1986, xxvi + 559 pp, $21.95 (P). [ISBN: 0-89588-249-3] Windows, menus, "quick-draw," memory management, file system,... . Uses Mac C to illustrate the role of tools from the Macintosh Toolbox to control the environment of a program. A useful supplement to Inside Macintosh, Apple's example-free guide to the Toolbox. C programs in the book are available on disk. LAS

Computer Programming, S(13-14). Computing Without Mathematics: Basic Pascal Applications. Jeffrey and Marvin Marcus. Computer Science Pr, 1986, xiii + 322 pp, $21.95 (P). [ISBN: 0-88175-105-7] Introduction to Basic and Pascal with examples emphasizing simple-minded text manipulation--as if students interested in word processing want (or need) to learn programming. LAS

Computer Programming, S(13-15). Using Turbo Pascal. Steve Wood. Osborne McGraw-Hill, 1986, ix + 304 pp, $19.95 (P). [ISBN: 0-07-881148-1] A useful introduction to the very popular Turbo Pascal, from installation to a full-scale application (a loan amortization program). Covers Turbo's predefined subprograms and suggests useful tools for an "include" library. (All code in the book is available on disk from the author.) Concludes with a quick reference guide. LAS

Software Systems, P. VAX/VMS Internals and Data Structures. Lawrence J. Kenah, Simon F. Bate. Digital Pr, 1984, xix + 795 pp, $45 (P). [ISBN: 0-932376-52-5] A standard reference which describes the VMS operating system for system managers and others who need to know truly intimate details of the executive. For Version 3.3, but most information also applies to 4.0 and later versions. Look in this unique work for the explicit steps VMS undertakes to create a subprocess, not for a description of the subprocess creation command (see VAX/VMS documentation set) or a definition and discussion of the concept of process (see any text on operating systems). Companion to Vax/VMS Guide to Writing a Device Driver. RB

Software Systems, P. Protext I. Ed: J.J.H. Miller. Boole Pr, 1984, vii + 248 pp, $65 (P). [ISBN: 0-906783-40-2] Proceedings of the first international conference on text processing held in Dublin, Ireland, during October 1984. Includes several papers on TeX, on interactive editing of mathematics, on UNIX editing tools, and on a host of more special systems. LAS

Computer Science, P. Office Automation: Concepts and Tools. Ed: Dionysios C. Tsichritzis. Topics in Inform. Syst. Springer-Verlag, 1985, xii + 441 pp, $29.50. [ISBN: 0-387-15129-X] A theoretical treatment applying such tools as the concepts of artificial intelligence, databases, programming languages, and machine architectures to a conceptual framework for information systems. JAS

Computer Science, P. Lecture Notes in Computer Science-201: Functional Programming Languages and Computer Architecture. Ed: Jean-Pierre Jouannaud. Springer-Verlag, 1985, vi + 413 pp, $22.80 (P). [ISBN: 0-387-15975-4] Proceedings from the conference held at Nancy, France, September 16-19, 1985. This conference was conceived of as a successor to the one held in Wentworth, New Hampshire in October 1981. JAS

Computer Science, P. Lecture Notes in Computer Science-196: Advances in Cryptology. Ed: G.R. Blakley, David Chaum. Springer-Verlag, 1985, ix + 491 pp, $25.10 (P). [ISBN: 0-387-15658-5] The proceedings of CRYPTO 84 which was held at the University of California, Santa Barbara on August 19-22, 1984. Contains about 40 papers. The focus is on the theory and application of cryptographic techniques. CEC

Computer Science, P. Semirings, Automata, Languages. Werner Kuich, Arto Salomaa. EATCS Mono. on Theor. Comp. Sci., V. 5. Springer-Verlag, 1986, 374 pp, $45. [ISBN: 0-387-13716-5] General framework for theory of nondeterministic, countable state automata using linear algebra and formal power series over semi-rings; ordinary (deterministic) language theory corresponding to Boolean semi-ring case. Matrix description of automata gives unified algebraic treatment for classical results. Behavior of finite automata (respectively PDA) expressed in terms of solutions to linear (respectively algebraic) equations. RM

Computer Science, T*(13-14: 2). Programming Principles Using Pascal. Roger H. Lamprey, Robert N. Macdonald, Morris W. Roberts. Harper & Row, 1985, xv + 672 pp. [ISBN: 0-06-043842-8] For courses CS1 and CS2 of ACM Curriculum '78. Organized around concepts rather than language features. Many fundamental algorithms and techniques. Three parts: an introduction to computer architecture and operating systems; the entire Pascal language; specific application areas (business data processing, numerical computation, systems programming, report-program generation). Examples are more advanced than in the usual beginning textbook. DFA

Computer Science, P. Lecture Notes in Computer Science-207: The Analysis of Concurrent Systems. Ed: B.T. Denvir, et al. Springer-Verlag, 1985, vii + 398 pp, $22.80 (P). [ISBN: 0-387-16047-7] Proceedings of a 1983 workshop and tutorial to find unifying framework for theories of concurrency, and inform academic researchers of industrial requirements. Overview of four approaches (algebraic, net theory, temporal logic, axiomatic) followed by a workshop to solve ten problems characteristic of real design problems. RM

Computer Science, P. 1985 Chapel Hill Conference on Very Large Scale Integration. Ed: Henry Fuchs. Computer Science Pr, 1985, xiii + 476 pp, $39.95. [ISBN: 0-88175-103-0] Twenty-five papers and a few abstracts from the 1985 Chapel Hill Conference. JAS

Computer Science, P. Database Machines. Ed: D.J. DeWitt, H. Boral. Springer-Verlag, 1985, 376 pp, $25 (P). [ISBN: 0-387-96200-X] Proceedings of the Fourth International Workshop held on Grand Bahama Island, March 1985. JAS

Computer Science, P. Lecture Notes in Computer Science-194: Automata, Languages and Programming. Ed: Wilfried Brauer. Springer-Verlag, 1985, ix + 520 pp, $27.40 (P). [ISBN: 0-387-15650-X] Proceedings of the twelfth colloquium on the title topic, held at Nafplion, Greece, July 15-19, 1985. JAS

Computer Science, S? Applied Concepts in Microcomputer Graphics. Bruce A. Artwick. Prentice-Hall, 1984, ix + 374 pp, $33.95. [ISBN: 0-13-039322-3] A rather technical book covering a wide variety of issues in both hardware and software. The "microcomputer" in the title is justified by its presentation of the Apple II memory map and quick survey of the 6845 registers and the memory maps for the IBM PC. Lots of technical details about display technology, matrix array processors, and details of raster scan standards in the U.S. and abroad, but no hint of homogeneous coordinates or GKS. Gives relatively few examples of algorithms and those few are given in BASIC. The contrasts between simple programs in BASIC and fairly detailed discussion of video "noise" and memory timing is considerable. JAS

Computer Science, P. Integrated Technology for Parallel Image Processing. Ed: S. Levialdi. Academic Pr, 1985, x + 236 pp, $39.50. [ISBN: 0-12-444820-8] A collection of papers from a workshop held at Polignano, Italy in June 1983. Topics include chip architecture, parallel processing, and system architecture. JAS

Computer Science, P. Fundamental Algorithms for Computer Graphics. Ed: Rae A. Earnshaw. NATO ASI Ser. F, V. 17. Springer-Verlag, 1985, xvi + 1042 pp, $59. [ISBN: 0-387-13920-6] Proceedings of an advanced study institute held at Ilkley, England, March 30 to April 12, 1985. JAS

Computer Science, P. Lecture Notes in Computer Science-208: Computation Theory. Ed: Andrzej Skowron. Springer-Verlag, 1985, vii + 397 pp, $22.80 (P). [ISBN: 0-387-16066-3] 32 papers from the fifth in a series of symposia organized by Humboldt University of Berlin and Warsaw University, this one held in Zaborów, Poland in December 1984. Deals with programming languages, complexity theory, computability theory, and artificial intelligence. LAS

Applications, S(15-17), P, L. Lecture Notes in Biomathematics-62: An Essay on the Importance of Being Nonlinear. Bruce J. West. Springer-Verlag, 1985, viii + 204 pp, $16.40 (P). [ISBN: 0-387-16038-8] A wide-ranging essay in natural philosophy emphasizing the intrinsically non-linear character of science--from the Gaussian distribution of error to strange attractors and chaotic phenomena. West argues that the robust character of linear models is misleading--that much of science actually depends on the structural stability of nonlinear phenomena. LAS

Applications, P. Viscoelasticity and Rheology. Ed: Arthur S. Lodge, Michael Renardy, John A. Nohel. Academic Pr, 1985, x + 445 pp, $35. [ISBN: 0-12-454940-3] Proceedings of an October 1984 conference at the Mathematics Research Center at the University of Wisconsin. LAS

Applications, P. Maritime Simulation. Ed: Moshe R. Heller. Springer-Verlag, 1985, xi + 290 pp, $29.50. [ISBN: 0-387-15620-8] Proceedings of a conference held in Munich in June 1985. Contains sections on a wide variety of topics including traffic control and fifth generation computer technology. JAS

Applications (Biology), P, L. Modeling Nature: Episodes in the History of Population Ecology. Sharon E. Kingsland. U of Chicago Pr, 1985, ix + 267 pp, $27.50. [ISBN: 0-226-43726-4] A scholarly history of population ecology, tracing especially the evolution of mathematical models through the work of Lotka, Volterra, Gause, Hutchinson, and MacArthur. An interesting case study in the gradual mathematization of a scientific discipline. LAS

Applications (Communication Theory), P. Spread Spectrum Communications. Marvin K. Simon, et al. Elect. Eng. Comm. & Signal Proc. Computer Science Pr, 1985. V. II, xix + 358 pp, $49.95 [ISBN: 0-88175-014-X]; V. III, xix + 423 pp. [ISBN: 0-88175-015-8] The history and theory of a recent and rapidly expanding electronic communications technique. JAS

Applications (Communication Theory), T(17-18: 1), P. Transform Coding of Images. R.J. Clarke. Academic Pr, 1985, xvi + 432 pp, $69.50. [ISBN: 0-12-175730-7] Digital image compression for storage and transmission at low bandwidth. Fast transform methods (producing spectral signals with most redundancy removed) made possible by advances in high-speed digital hardware. Statistical in nature (no feature abstraction) with model of human visual system to improve performance; consideration of hybrids with predictive methods. RM

Applications (Economics), P. Lecture Notes in Economics and Mathematical Systems-251: Input-Output Modeling. Ed: A. Smyshlyaev. Springer-Verlag, 1985, vi + 259 pp, $20.50 (P). [ISBN: 0-387-15698-4] Proceedings of the fifth meeting of a task force of the international institute for applied systems analysis (IIASA) held at Laxenburg, Austria, in October 1984. LAS

Applications (Engineering), T(16-17: 1), P. Introduction to Random Processes With Applications to Signals and Systems. William A. Gardner. Macmillan, 1986, xiii + 434 pp, $32.50. [ISBN: 0-02-948790-0] The emphasis is on random noise in an engineering environment. Begins with a review of relevant probability topics, then introduces the standard models: Weiner, Poisson, Markov, followed by discussions of ergodicity, duality, filters, spectral density, cyclostationary processes and

error estimation. Extensive exercises and bibliography. TAV

Applications (Engineering), P. Computer-Aided Processes in Instruction and Research. Ed: George C. Beakley, C.R. Haden. Academic Pr, 1985, xi + 373 pp, $29.50. [ISBN: 0-12-083520-7] Two dozen papers selected to illustrate a wide range of applications of computers in engineering practice and instruction. LAS

Applications (Engineering), S, P, L. Local Area Networks: An Introduction to the Technology. John E. McNamara. DEC, 1985, viii + 165 pp, $28 (P). [ISBN: 0-932376-79-7] An informative, well-organized introduction to vocabulary, concepts, hardware, and issues involved with local area networks (LANS). A good primer on the subject. LAS

Applications (Epidemiology), S(16-17), L. Lecture Notes in Biomathematics-59: Hybrid Models of Tropical Infections. Ingemar Nåsell. Springer-Verlag, 1985, vi + 206 pp, $16.40 (P). [ISBN: 0-387-15978-9] Models with both stochastic and deterministic ingredients. Assumes some familiarity with Markov chains in continuous time and discrete state space, and with qualitative theory of ordinary differential equations. Formulates and studies classical models for malaria and schistosomiasis and extensions of these models. DFA

Applications (Physics), S*, P*, L.** QED: The Strange Theory of Light and Matter. Richard P. Feynman. Princeton U Pr, 1985, xi + 158 pp, $18.50. [ISBN: 0-691-08388-6] A tour de force of scientific popularization--an intellectually honest lay account of quantum electrodynamics based on four public lectures at UCLA. Using extensive arrow diagrams to represent probabilities, but no equations, Nobel laureate Feynman shows convincingly that light does not always travel in straight lines or precisely at the speed of light. His simple yet honest account renders this arcane theory in the plainest language possible. LAS

Applications (Physics), P. Superstrings: The First 15 Years of Superstring Theory. Ed: John H. Schwarz. World Scientific, 1985, $42 set (P); $90 set [ISBN: 9971-978-67-9; 9971-978-66-0]. Volume 1, xi + 574 pp; Volume 2, ix + 564 pp. Reprints of 55 source papers in the theory of superstrings, the currently fashionable "theory of everything" that incorporates foundational ideas in all parts of physics. An invaluable resource for anyone who would learn about superstrings. LAS

Applications (Physics), P. Lecture Notes in Mathematics-1159: Schrödinger Operators. Ed: S. Graffi. Springer-Verlag, 1985, viii + 272 pp, $17.60 (P). [ISBN: 0-387-16035-3] Six lectures on topics in functional analysis, partial differential equations, and quantum mechanics related to Schrödinger operators. From a 1984 C.I.M.E. Session in Como, Italy. LAS

Applications (Physics), P. Spherical Harmonics and Tensors for Classical Field Theory. M.N. Jones. Appl. & Eng. Math. Ser., V. 2. Wiley, 1985, xiii + 230 pp, $45. [ISBN: 0-471-90766-9] Applications of group theory and group representations to problems in classical field theory. Spherical symmetry is usually assumed; geophysical and other applications follow. The basics of group theory and group representations are covered. PZ

Applications (Physics), P. Stochastic Processes Applied to Physics and other Related Fields. Ed: B. Gomez, et al. World Scientific, 1983, xxv + 781 pp, $75. [ISBN: 9971-950-56-1] Lecture series from the 1982 Latin America summer school of physics held at Cali, Colombia: seven courses on the role of stochastic processes in each of various subfields of physics. LAS

Applications (Physics), P. Lecture Notes in Physics-224: Supernovae as Distance Indicators. Ed: Norbert Bartel. Springer-Verlag, 1985, vi + 226 pp, $13.70 (P). [ISBN: 0-387-15206-7] Proceedings of a workshop held at the Harvard-Smithsonian Center for Astrophysics, September 1984. Organized into four sections: review; observations of supernovae and supernova remnants; radio supernovae and inference of extragalactic distances; optical supernovae and inference of extragalactic distances; and the after-dinner talk by B.G. Marsden, "Supernovae Up Close." Virtually devoid of mathematics. MU

Applications (Physics), P. Lecture Notes in Physics-227: Computer Methods and Borel Summability Applied to Feigenbaum's Equation. Jean-Pierre Eckmann, Peter Wittwer. Springer-Verlag, 1985, xiv + 297 pp, $16.60 (P). [ISBN: 0-387-15215-6] The main purpose of this work is to prove the existence of a class of solutions of Feigenbaum's functional equation. The main themes are Feigenbaum universality, Ecalle's theory of resurgent functions, constructive aspects of Borel summation, and techniques for computer assisted proofs. CEC

Reviewers

DFA: David F. Appleyard, Carleton; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; RD: Roger Day, St. Olaf; JD-B: John Dyer-Bennet, Carleton; GF: Giovanna Fjelstad, St. Olaf; SG: Steven Galovich, Carleton; BH: Bruce Hanson, St. Olaf; PH: Paul Humke, St. Olaf; KK: Kenneth Kaminsky, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; AM: Alan Magnuson, St. Olaf; SM: Steve McKelvey, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MT: Michael Tveite, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

BSB^T and $(B^{-1})^T CB^{-1}$ are symmetric. Note that the first factor BSB^T is nonsingular.

THEOREM 2. *Every real square matrix A is a product of two real symmetric matrices.*

Proof. Let $A = BJB^{-1}$ as in Proposition 1 and $J = SC$ as in Proposition 2. Partition the matrices B, J and S all in the same way with respect to the columns:

$$B = \left(B_{(1)} B_{(2)} B_{(3)}\right); \quad J = \begin{bmatrix} J_{(1)} & & 0 \\ & J_{(2)} & \\ 0 & & J_{(3)} \end{bmatrix}; \quad S = \begin{bmatrix} S_{(1)} & & 0 \\ & S_{(2)} & \\ 0 & & S_{(3)} \end{bmatrix}.$$

A is real, hence J is balanced (see Corollary). By permutation of the columns of B we can see to it that $J_{(2)} = \bar{J}_{(1)}$ (where the λ's in $J_{(1)}$ are all different from the λ's in $\bar{J}_{(1)}$) and that $J_{(3)}$ contains *all* real blocks of J (so $J_{(3)}$ is real and $S_{(2)} = \bar{S}_{(1)} = S_{(1)}$). Of course $J_{(1)}$ or $J_{(3)}$ can be "empty". $AB = BJ$ in partitioned form gives:

$$\left(AB_{(1)}\ AB_{(2)}\ AB_{(3)}\right) = \left(B_{(1)}J_{(1)}\ B_{(2)}\bar{J}_{(1)}\ B_{(3)}J_{(3)}\right).$$

$AB_{(1)} = B_{(1)}J_{(1)}$, hence $A\bar{B}_{(1)} = \bar{B}_{(1)}\bar{J}_{(1)}$; $AB_{(3)} = B_{(3)}J_{(3)}$, hence $A\bar{B}_{(3)} = \bar{B}_{(3)}J_{(3)}$ ($B_{(3)}$ is of full rank). This means that there exists a *real* matrix $D_{(3)}$ of full rank such that $AD_{(3)} = D_{(3)}J_{(3)}$. Define $D = (B_{(1)}\bar{B}_{(1)}D_{(3)})$; then $AD = DJ$. D is nonsingular: $B_{(1)}$ (so $\bar{B}_{(1)}$) and $D_{(3)}$ are of full rank. The columns of $B_{(1)}$ and $\bar{B}_{(1)}$, and the columns of $B_{(1)}(\bar{B}_{(1)})$ and $D_{(3)}$ are mutually *independent*, being (generalized) eigenvectors associated with *different* eigenvalues. Hence D is nonsingular. Now

$$A = DJD^{-1} = DSCD^{-1} = [DSD^T]\left[(D^{-1})^T CD^{-1}\right].$$

Both factors are symmetric and DSD^T is, as a product of nonsingular matrices, nonsingular. In partitioned form we have:

$$DSD^T = \left(D_{(1)}S_{(1)}D_{(1)}^T + \bar{D}_{(1)}S_{(1)}\bar{D}_{(1)}^T\right) + D_{(3)}S_{(3)}D_{(3)}^T$$

which is, as a sum of two real matrices, also real.
$(D^{-1})^T CD^{-1} = (DSD^T)^{-1}A$ is, as a product of real matrices, also real.

References

1. D. H. Carlson, On real eigenvalues of complex matrices, Pacific J. Math., 15 (1965) 1119.
2. G. Frobenius, Ueber die mit einer Matrix vertauschbaren Matrizen, Sitzungsber. Preuss. Akad. für Wiss. (1910) 3.
3. Ben Noble and J. W. Daniel, Applied Linear Algebra, Prentice-Hall, 1977.
4. Chi Song Wong, Characterization of products of symmetric matrices, Linear Algebra and Appl., 42 (1982) 243.
5. O. Taussky, The role of symmetric matrices in the study of general matrices, Linear Algebra and Appl., 5 (1972) 147.

HOW MANY ZEROS DOES A CONTINUOUS FUNCTION HAVE?

TOMÁS DOMÍNGUEZ BENAVIDES
Facultad de Matemáticas, c/Tarfia s/n, Sevilla-12, Spain

Let $I = [a, b]$ and let $\mathscr{C}_O(I)$ be the space of those continuous real-valued functions on I having at least one zero with the topology of uniform convergence. We denote by d the metric in $\mathscr{C}_0(I)$,

$$d(f, g) = \max\{|f(x) - g(x)|: x \in I\}.$$

If $f \in \mathscr{C}_0(I)$, let $Z(f)$ be the set of zeros of f.

We use Baire category as an indication of the "largeness" of a subset of $\mathscr{C}_0(I)$. Recall that a set of the first category is usually thought to be "small"; it is in fact often called *meager*. The complement of such a set is thought to be "large"; it is often called *residual*.

We use cardinality and Lebesgue measure m to measure the "largeness" of subsets of I.

Our basic result is that $Z(f)$ is "in general" *small* in the sense of the Lebesgue measure, but *large* in the sense of cardinality. Such results are usually called "generic results". A specialized study about such results can be found in [**2**] and the references given there. Specifically, we prove the following:

THEOREM 1. *There is a residual subset M of $\mathscr{C}_0(I)$ such that for every $f \in M$,*

$$m(Z(f)) = 0.$$

THEOREM 2. *There is a residual subset N of $\mathscr{C}_0(I)$ such that for every $f \in N$,*

$$\operatorname{card}(Z(f)) = c.$$

For the proof of Theorem 1 we use the following observation, which can be easily proved by using the Weierstrass approximation Theorem.

LEMMA 1. *The set $\{p \in \mathscr{C}_0(I)\colon p \neq 0;\ p$ is a polynomial$\}$ is a dense subset of $\mathscr{C}_0(I)$.*

Proof of Theorem 1. Define

$$M_n = \{f \in \mathscr{C}_0(I) : m(Z(f)) < 1/n\}.$$

We claim that M_n is an open set. Indeed, let f be a function in M_n. Since $m(Z(f)) < 1/n$, there exists an open set G such that $Z(f) \subset G$, and $m(G) < 1/n$. Let η be

$$\inf\{|f(x)|\colon x \in I \setminus G\}.$$

We have $\eta > 0$, because otherwise there is sequence $\{x_n\}$ in $I \setminus G$ such that $\{|f(x_n)|\} \to 0$. Since $I \setminus G$ is a compact set, we can assume that $\{x_n\}$ converges to a point $x \in I \setminus G$. Thus $f(x) = 0$, contradicting $x \notin Z(f)$. Let g be a function in $\mathscr{C}_0(I)$ such that $d(f, g) < \eta$. If $x \in Z(g)$, we have

$$|f(x)| = |f(x) - g(x)| < \eta.$$

Thus x belongs to G. Hence $m(Z(g)) \leqslant m(G) < 1/n$, which implies $g \in M_n$. Define

$$M = \bigcap_{n=1}^{\infty} M_n.$$

Since M contains the polynomials $p \in \mathscr{C}_0(I)$, $p \neq 0$, we have from Lemma 1 that M is a dense G_δ. Thus M is a residual subset of $\mathscr{C}_0(I)$. It is clear that

$$M = \{f \in \mathscr{C}_0(I) : m(Z(f)) = 0\}.$$

Proof of Theorem 2. Let $A = \{f \in \mathscr{C}_0(I)\colon \operatorname{card} Z(f) < c\}$. We shall prove that A is of Baire first category in $\mathscr{C}_0(I)$. Denote $\mathscr{F} = \{I_n\colon n \in \mathbb{N}\}$ the collection of those closed subintervals of I whose extremes are rational numbers. Define

$$A_n = \{f \in A\colon f \text{ has a unique zero in } I_n\}.$$

We claim that A_n is nowhere dense; the interior of its closure is empty. Assume by contradiction that there exists $f \in A_n$ and $\varepsilon > 0$ such that $B(f, \varepsilon) \subset \operatorname{clos} A_n$. Let $\xi \in I_n$ be the (unique) zero of f in I_n. Choose $\delta > 0$, $\delta < \varepsilon/2$, such that $|f(x) - f(\xi)| < \varepsilon/2$ whenever $|x - \xi| < 2\delta$. Define a continuous function $g\colon I \to \mathbb{R}$, $d(f, g) < \varepsilon$, such that

$$g(x) = \begin{cases} (x - \xi)\sin|x - \xi|^{-1}, & \text{if } 0 < |x - \xi| < \delta, \\ f(x), & \text{if } |x - \xi| > 2\delta. \end{cases}$$

Write $I_n = [a_n, b_n]$ and assume that $\xi \neq a_n$, $a_n < \xi - \delta$ (if $\xi \neq b_n$ a similar argument can be used). Denote by x_1, x_2, x_3 the three first zeros of $\cos(1/x)$ in the interval $(-\delta, 0)$ and let η be $|x_3|/2$. Let h be a function in $B(g, \eta)$. Since

$$g(x_k + \xi) = \pm x_k \quad (k = 1,2,3),$$

we have that sign $h(x_k + \xi) =$ sign $g(x_k + \xi)$. Thus

$$\operatorname{sign} h(x_k + \xi) \neq \operatorname{sign} h(x_{k+1} + \xi) \quad (k = 1,2)$$

because sign $g(x_k + \xi) \neq$ sign $g(x_{k+1} + \xi)$. Therefore there exists a zero of h in the interval $(x_k + \xi, x_{k+1} + \xi)$. Then h has at least two zeros in I_n. This implies

$$B(g, \eta) \cap A_n = \varnothing,$$

contradicting $g \in B(f, \varepsilon) \subset \operatorname{clos} A_n$.

Let f be a function in A. The set $Z(f)$ has some isolated point because every closed subset of I without isolated points has cardinality c [**1**, Th. 6.65]. Let ξ be an isolated zero of f. There exists $I_n \in \mathscr{F}$ such that ξ is the unique zero of f in I_n. Then f belongs to A_n. Thus the set $A \subset \bigcap_{n=1}^{\infty} A_n$ is of Baire first category.

References

1. E. Hewitt and K. Stromberg, Real and Abstract Analysis, Springer, Berlin, 1965.
2. J. Myjak, Orlicz type category theorems for functional and differential equations, Dissertations Math., vol. CCVI.

QUADRATURE WITH GENERALIZED MEANS

JET WIMP
Department of Mathematical Sciences, Drexel University, Philadelphia, PA 19104

The famous trapezoidal rule for numerical integration over a single interval is

$$\int_a^b f(x)\,dx \cong (b-a)[f(a) + f(b)]/2, \tag{1}$$

and when the domain of f is divided into n equal intervals and the rule applied to each interval the result is the composite trapezoidal rule,

$$\int_a^b f(x)\,dx \cong [(b-a)/n] \sum_{k=0}^{n}{}'' f(x_k), \tag{2}$$

$$x_k = a + k(b-a)/n,$$

where the double prime on the summation sign indicates that the first and the last term in the sum are to be halved.

It is easily shown that when f is a linear function, $f = \alpha x + \beta$, the above formulas are exact, i.e., the " $\cong$ " may be replaced by " $=$ ". The trapezoidal rule is sometimes considered the poor cousin in numerical analysis, but it can be surprisingly effective, often giving better results than other more sophisticated schemes for numerical integration, particularly when the function f lacks smoothness.

The rule (1) can be considered the result of approximating the integral on the left by the arithmetic mean of two values of the function f. However, means other than the arithmetic mean arise frequently in applied mathematics, for instance, the geometric mean, $M(x, y) = (xy)^{1/2}$ or the harmonic mean, $M(x, y) = (x^{-1} + y^{-1})^{-1}$. One question is, do these and other means also give rise to numerical integration formulas? The answer, as we shall see, is yes, and furthermore composite rules for these formulas analogous to (2) are readily defined and they are as accurate as the composite trapezoidal rule. It turns out that they are exact, when applied to a single

A COLORING PROOF OF A GENERALISATION OF FERMAT'S LITTLE THEOREM

C. J. SMYTH
Department of Electronic Systems Engineering, University of Essex, Colchester CO4 3SQ, United Kingdom

Before stating the main result of this note, I ask the reader to fill in the missing formula in the following table:

	$(a, n) = 1$	all integers a
$n = p$ prime	$a^{p-1} \equiv 1 \pmod p$	$a^p \equiv a \pmod p$
n composite	$a^{\phi(n)} \equiv 1 \pmod n$	?

(Here ϕ is Euler's totient function: $\phi(n) = n\prod_{p|n}(1 - \frac{1}{p})$).

The answer to this question seems little-known to mathematicians, even to number theorists. The reason for this seems to be its non-appearance in most of the standard reference books. The missing result is a beautiful one:

$$\sum_{d|n} a^d \mu\left(\frac{n}{d}\right) \equiv 0 \pmod n \tag{1}$$

(μ being the Möbius function defined by $\sum_{n=1}^{\infty} \frac{\mu(n)}{n^s} = \prod_{p \text{ prime}} (1 - p^{-s})$). Its history is chronicled in Dickson [**1**, pp. 82–86]. Gauss' proof of the result, but only for a prime, was published posthumously in 1863. It was not until 1880–83 that four independent proofs for all a were published by Kantor, Weyr, Lucas and Pellet (for precise references see [**1**]; see also [**5**]).

Let $\theta_1, \theta_2, \ldots, \theta_D$ be the zeros of a monic polynomial with integer coefficients, and put

$$S_n = \theta_1^n + \theta_2^n + \cdots + \theta_D^n. \tag{2}$$

Then the result

$$S_p \equiv S_1 \pmod p \tag{3}$$

proved by Schönemann [**4**] in 1839 is a generalisation of

$$a^p \equiv a \pmod p \tag{4}$$

to which, of course, (3) reduces if $D = 1$.

The result I will prove is the following:

$$\sum_{d|n} S_d \mu\left(\frac{n}{d}\right) \equiv 0 \pmod n \tag{5}$$

generalising both (1) and (3).

In 1872 Petersen [**3**] proved Fermat's Little Theorem (4) by the following argument (for $a > 0$): Suppose one has p boxes, arranged in a circle, to be colored with a colors. There are a^p colorings in all, and a colorings with every box the same color. The $a^p - a$ remaining colorings can be arranged in sets of p, since the p rotations of any one of these colorings are all distinct. Hence $p|(a^p - a)$.

Thue [**6**] in 1910 published a proof of (1) by generalising this idea. (His proof is neatly summarised in [**1**, p. 82]. Thue uses (1) to prove $a^{\phi(n)} \equiv 1 \pmod n$ for $(a, n) = 1$.) Here we will prove (5), by generalising a bit further. Our result is:

THEOREM. *Let $a_1, a_2, a_3, \ldots$ be an infinite sequence of non-negative integers. Suppose we are given n boxes, arranged in a circle. In some of the spaces between adjacent boxes, a barrier or*

partition is placed, the number of such partitions ranging from 1 *to* n. *Suppose that a group of* j *boxes between two partitions must be painted the same color, from a palette of* a_j *different colors. Then the total number* s_n^* *of such "partition-colorings" of the boxes satisfies*

$$(6^*)\qquad s_n^* = a_1 s_{n-1}^* + a_2 s_{n-2}^* + \cdots + a_{n-1}s_1^* + na_n \quad (n = 1,2,\ldots),$$

$$(5^*)\qquad \sum_{d|n} s_d^* \mu\left(\frac{n}{d}\right) \equiv 0 (\bmod n).$$

Note that the coloring does not always specify the partition: when there is only one partition, all n boxes are colored alike, in a_n possible colors, but we count $n\,a_n$ total partition-colorings, taking into account the n possible positions for the partition.

Before proving the theorem, we show how (5) follows from it. Let

$$\begin{aligned} P(x) &= (x-\theta_1)(x-\theta_2)\cdots(x-\theta_D) \\ &= x^D - a_1x^{D-1} - a_2x^{D-2} - \cdots - a_{D-1}x - a_D, \end{aligned}$$

say. Then, as Newton showed, the S_n defined by (2) satisfy

$$(6)\qquad S_n = a_1S_{n-1} + a_2S_{n-2} + \cdots + a_{n-1}S_1 + na_n \quad (n = 1,2,\ldots),$$

where we put $a_k = 0$ for $k > D$. This follows straight from the fact that

$$D + \sum_{n=1}^{\infty} S_n z^n = \sum_{j=1}^{D} \frac{1}{1-\theta_j z} = P'(z^{-1})/(zP(z^{-1})).$$

Since $\{S_n\}$ is uniquely specified by (6), we have from the Theorem that $S_n = s_n^*$ and so (5) holds, when $a_1,\ldots,a_D$ are non-negative.

If any a_i in (7) are negative, we can still verify (5), as follows: fix n, and put

$$a_i = a_i^+ - k_i n \quad (i = 1,\ldots,D),$$

where the a_i^+ are positive, and the k_i integers. Define $S_1^+, S_2^+,\ldots,S_n^+$ by

$$(6^+)\qquad S_j^+ = a_1^+S_{j-1}^+ + \cdots + a_{j-1}^+S_1^+ + ja_j^+ \quad (j = 1,\ldots,n).$$

Then $\sum_{d|n} S_d^+\mu(\frac{n}{d}) \equiv 0 (\bmod n)$. But $S_j \equiv S_j^+ \pmod n$ for $j = 1,\ldots,n$ by (6^+) and so (5) follows for arbitrary integers a_i.

Proof of the theorem. Label the boxes $B_0, B_2,\ldots,B_{n-1}$, going around clockwise. Suppose, for a particular partition-coloring that, starting from B_0, the first complete group of boxes between partitions goes from B_t to $B_{t+u-1 (\bmod n)}$. Then by removing this group of u boxes, and closing up the gap (leaving a partition there), we obtain an associated partition-coloring of $n-u$ boxes, if $u<n$. (If $u=n$ we know already that we obtain $n\,a_n$ possible colorings.)

Conversely, any partition-coloring of $n-u$ boxes, $u<n$, can be used to construct a partition-coloring of n boxes, by inserting a group of u boxes, followed by a partition, immediately after the first partition found, on proceeding from B_0 clockwise. Thus this correspondence is 1–1, and, since the inserted group of u boxes can be colored in a_u ways,

$$s_n^* = \sum_{u=1}^{n-1} a_u s_{n-u}^* + na_n$$

which is (6^*).

We now show that there are integers r_n divisible by n such that

$$(7)\qquad s_n^* = \sum_{d|n} r_d,$$

from which $n|r_n = \sum_{d|n} s_d^* \mu(\frac{n}{d})$, and hence (5*) follows immediately by Möbius inversion (see, e.g., [2, p. 234]).

We define r_n to be the number of partition-colorings of n boxes which are distinct from all their rotations. Then clearly $n|r_n$. Further, the partition-colorings of n boxes which have the property that rotation by d places, but not fewer, produces the same coloring are obtained in the following way: Take any of the r_d colorings of d boxes referred to above, and repeat the pattern n/d times, with a partition between each pattern, to obtain a coloring of n boxes. There are r_d such colorings, and so the total number s_n^* colorings is given by (7). This completes the proof of the theorem.

I would like to thank Professor David Boyd for useful discussions concerning this topic, and Professor A. Schinzel for providing reference [5]. This note was written while the author was at James Cook University, Townsville, Australia.

References

1. L. E. Dickson, History of the Theory of Numbers, Vol. 1, Chelsea, New York, 1971.
2. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, 4th ed., Oxford University Press, 1960.
3. J. Petersen, Tidsskrift for Mathematik (3), 2, 1972, 64–65 (Danish).
4. T. Schönemann, J. Für Math. (Crelle), 19 (1839), p. 290, 31 (1846), p. 288.
5. T. Szele, Une généralisation de la congruence de Fermat, Mat. Tidsskr, B. 1948, 57–59 (1948).
6. A. Thue, Ein Kombinatorischer Beweis eines Satzes von Fermat, in Selected Mathematical Papers of Axel Thue, Universitelsforlaget, 1977 (originally Kra. Vidensk. Selsk, Skrifter. I. Mat. Nat. Kl. 1910, No. 3).

SOME REMARKS ON FUNCTIONS WITH ONE-SIDED DERIVATIVES

A. D. MILLER* AND R. VÝBORNÝ
University of Queensland, St. Lucia, 4067 Queensland, Australia

An important theorem in introductory calculus relates the monotonicity of a function to the sign of its derivative: If f is continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b), then f is monotonic increasing (decreasing) if and only if $f' \geqslant 0$ ($\leqslant 0$) on (a, b). (A simple corollary to this result states that f is constant on $[a, b]$ if and only if $f' = 0$ on (a, b).) Most standard proofs of the sufficiency part of this theorem use the classical Mean Value Theorem of Calculus.

Despite the simple geometrical interpretation of the Mean Value Theorem, namely, the chord of the graph of f must be parallel to the tangent at some intermediate point, experience in teaching elementary calculus courses shows that whereas students find the above monotonicity theorem geometrically plausible, they often have difficulty in grasping the meaning of the Mean Value Theorem. It is interesting to note that the usual textbook proof of the Mean Value Theorem is due to Bonnet, and first appeared in print in 1868. For a detailed historical account of the Mean Value Theorem see [2]. Surprisingly, it does not seem generally known that a few years later Scheeffer [6] in an investigation of the uniqueness of the anti-derivative introduced an idea which is able to provide an alternative, simple and intuitive approach to the above monotonicity result. In this note we wish to revive this idea. However, our main aim is to prove the monotonicity result but requiring only conditions on one-sided derivatives. In doing this we refine the remark of Knight [5].

THEOREM 1. *Let f be a continuous function on $[a, b]$. If for each $x \in (a, b)$ one of the one-sided*

*Now at Centre for Mathematical Analysis, Australian National University, Canberra, A.C.T. 2600, Australia.

derivatives $f'_+(x)$ *or* $f'_-(x)$ *exists, and is nonnegative (possibly* $+\infty$), *then* f *is monotonic increasing.*

[By also applying this result to $-f$, we see that if for each $x \in (a, b)$ either $f'_+(x) = 0$ or $f'_-(x) = 0$, then f is constant.]

Proof of Theorem 1. Suppose f is not monotonic increasing. We may then find points $\alpha, \beta \in [a, b]$ with $\alpha < \beta$ but $f(\alpha) > f(\beta)$. The proof proceeds by locating a point in (α, β) at which both of the one-sided derivatives must be negative (if they exist). We do this by considering the last intersection of the graph of f with an appropriate straight line L_t (see Fig. 1). Analytically, consider the linear function $L_t(x) = -t(x - \alpha) + c$, where, say,

$$c = \frac{1}{2}(f(\alpha) + f(\beta))$$

and $t > 0$ has been chosen so that $L_t(\beta) > f(\beta)$; e.g., choose t so as to satisfy

$$0 < t < \frac{f(\alpha) - f(\beta)}{2(\beta - \alpha)}.$$

Let $\xi_t = \sup\{x \in [\alpha, \beta]: f(x) \geqslant L_t(x)\}$. A standard argument using the continuity of f shows that $\xi_t \in (\alpha, \beta)$, and furthermore that $f(\xi_t) = L_t(\xi_t)$.

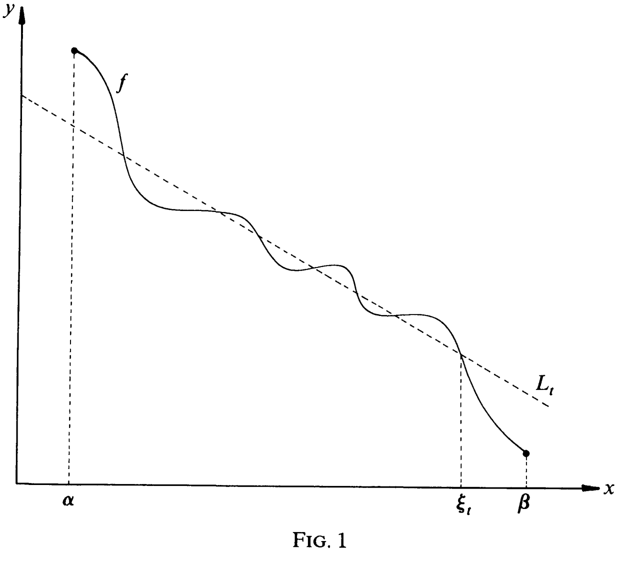


FIG. 1

By the definition of ξ_t we see that

(i) For all $x \in (\xi_t, \beta)$, $f(x) < L_t(x)$ and hence

$$\frac{f(x) - f(\xi_t)}{x - \xi_t} \leqslant \frac{L_t(x) - L_t(\xi_t)}{x - \xi_t} = -t,$$

which shows that provided the right-hand derivative exists at ξ_t,

(1) $$f'_+(\xi_t) < 0.$$

(ii) There is an increasing sequence $x_n \uparrow \xi_t$ such that $f(x_n) \geqslant L_t(x_n)$, and hence

$$\frac{f(\xi_t) - f(x_n)}{\xi_t - x_n} \leqslant \frac{L_t(\xi_t) - L_t(x_n)}{\xi_t - x_n} = -t,$$

showing that as long as the left-hand derivative exists at ξ_t,

$$f'_-(\xi_t) < 0. \tag{2}$$

The inequalities (1) and (2) contradict the assumption that one of the one-sided derivatives of f exists at ξ_t and is nonnegative. Thus the proof is complete.

The classical mean value theorem of differential calculus states that for a real valued function f continuous on the closed interval $[a, b]$, and differentiable on the open interval (a, b), there is an intermediate point $c \in (a, b)$ such that

$$f(b) - f(a) = (b - a)f'(c). \tag{3}$$

Often in practice the existence of the point c is not in itself of importance, but rather that (3) leads to the estimate

$$\inf\{f'(x): x \in (a, b)\} \leqslant \frac{f(b) - f(a)}{(b - a)} \leqslant \sup\{f'(x): x \in (a, b)\}.$$

Our Theorem 1 permits a generalization of this estimate.

COROLLARY. *Suppose f is continuous on $[a, b]$. For each $x \in (a, b)$ assume that one of the one-sided derivatives $f'_+(x)$ or $f'_-(x)$ exists, and denote it by $f'_*(x)$. (Thus f'_* may at some points be the left-hand derivative, and at other points the right-hand derivative.) Then*

$$\inf\{f'_*(x): x \in (a, b)\} \leqslant \frac{f(b) - f(a)}{(b - a)} \leqslant \sup\{f'_*(x): x \in (a, b)\}. \tag{4}$$

Proof. Suppose $m = \inf\{f'_*(x): x \in (a, b)\}$. Notice that $m \neq \infty$, while if $m = -\infty$ the left-hand estimate of (4) is trivial; so we need only consider the case m finite. Define

$$F(x) = f(x) - m(x - a) \quad \text{for all } x \in [a, b].$$

Theorem 1 can now be applied to F to obtain the inequality

$$F(b) \geqslant F(a),$$

that is

$$f(b) - m(b - a) \geqslant f(a).$$

But this immediately yields the left-hand side of (4). The right-hand side of (4) follows similarly.

The proof of Theorem 1 in fact establishes a stronger result. If E is a countable set, then ξ_t can always be chosen to lie outside E (since t may range over a continuum, and $\xi_t \neq \xi_s$ for $t \neq s$). Moreover, (1) and (2) can be replaced by the inequalities $D^+f(\xi_t) < 0$ and $D_-f(\xi_t) < 0$, respectively, where $D^+f(\xi_t)$ and $D_-f(\xi_t)$ are, respectively, the upper right and lower left Dini derivates of f at ξ_t. (Recall that for a real valued function the Dini derivatives are defined by

$$D^+f(x) = \limsup_{h \downarrow 0} \frac{f(x + h) - f(x)}{h},$$

$$D_+f(x) = \liminf_{h \downarrow 0} \frac{f(x + h) - f(x)}{h},$$

$$D^-f(x) = \limsup_{h \uparrow 0} \frac{f(x + h) - f(x)}{h},$$

$$D_-f(x) = \liminf_{h \uparrow 0} \frac{f(x + h) - f(x)}{h}.$$

These always exist, though possibly with values $\pm\infty$.) Hence Theorem 1 remains valid if for the second sentence of the statement we substitute: "If for each $x \in (a, b) - E$, either $D^+f(x) \geqslant 0$

or $D_-f(x) \geqslant 0$ (possibly $+\infty$), then f is monotonic increasing". (By a simple modification of the above proof it follows that the requirement "either $D^+f \geqslant 0$ or $D_-f \geqslant 0$", could just as well read "either $D^-f \geqslant 0$ or $D_+f \geqslant 0$".) The corresponding condition guaranteeing the constancy of f is that for each $x \in (a, b) - E$ either $D^+f(x) = 0$ or $D_-f(x) = 0$ (or for each $x \in (a, b) - E$ either $D^-f(x) = 0$ or $D_+f(x) = 0$).

We may quite naturally wonder to what extent this extended version of Theorem 1 is the weakest possible characterization of monotonicity in terms of the Dini derivates. The negative of the Cantor function [**3**;p. 96] shows that only requiring the exceptional set E to be of measure zero is not sufficient to ensure that a function is monotonic increasing. A further refinement of Theorem 1 is given by the following.

THEOREM 2. *Let f be continuous on $[a, b]$, and suppose $E \subseteq (a, b)$ is of measure zero. Assume that for each $x \in (a, b) - E$, either $D^+f(x) \geqslant 0$ or $D_-f(x) \geqslant 0$ (possibly $+\infty$). If in addition at each point of E, except perhaps those from some countable set, one of D^+f or D_-f is not $-\infty$, then f is monotonic increasing.*

Proof of Theorem 2. It is sufficient to construct an increasing continuous function ϕ where $\phi'(x) = +\infty$ on E. The proof is then completed by applying the strengthened version of Theorem 1 to $f + \varepsilon\phi$, and then letting $\varepsilon \to 0$.

There are a number of methods of constructing such a ϕ. We follow § 3.1.2 of [**7**].

Recall that if $E \subset (a, b)$ is a set of measure zero then there is a countable set of open intervals (a_i, b_i) $i = 1,2,\ldots,$ such that

(i) $E \subset \bigcup_{i=1}^{\infty} (a_i, b_i) \subset (a, b)$,

(ii) $\sum_{i=1}^{\infty} (b_i - a_i) < \infty$,

(iii) $x \in E \Rightarrow x \in (a_i, b_i)$ for infinitely many i.

Define the function

$$\phi_i(x) = \begin{cases} 0, & a \leqslant x < a_i, \\ x - a_i, & a_i \leqslant x \leqslant b_i, \\ b_i - a_i, & b_i < x \leqslant b, \end{cases}$$

and consider the series

$$\phi(x) = \sum_{i=1}^{\infty} \phi_i(x).$$

Clearly each ϕ_i is continuous and increasing on $[a, b]$, and since the series is uniformly convergent (by (ii)) the sum $\phi(x)$ is also continuous and increasing on $[a, b]$.

Let $x \in E$. By (iii) the number $x \in (a_i, b_i)$ for infinitely many indices i. Let N be any natural number. Choose N of the intervals (a_i, b_i) containing x and denote their intersection by I. If $y \in I$, then

$$\frac{\phi_i(x) - \phi_i(y)}{x - y} = 1$$

for N indices i. Consequently

$$\frac{\phi(x) - \phi(y)}{x - y} = \sum_i \frac{\phi_i(x) - \phi_i(y)}{x - y} \geqslant N.$$

But since N is arbitrary we conclude that $\phi'(x) = \infty$.

A special case of Theorem 2 is worth mentioning as a corollary.

COROLLARY. *Let f be continuous on $[a, b]$, and suppose $E \subset (a, b)$ is of measure zero. Assume that for each $x \in (a, b) - E$, $D^+f(x) \geqslant 0$ (possibly $+\infty$). If in addition at each point of E, except perhaps those from some countable set, D^+f is not $-\infty$, then f is monotonic increasing.*

This corollary can be found in the literature (see § 7.2.2.1 in [**1**]). However, the existing proofs are more involved than our proof of Theorem 2. It is interesting to note that Theorem 2 can be recovered from the above corollary by using a non-trivial result of Grace Young [**8**] which states that $D^+f \geqslant D_-f$ except perhaps on a countable set.

The hypothesis "either $D^+f(x) \geqslant 0$ or $D_-f(x) \geqslant 0$" in both Theorem 1 and Theorem 2 cannot be weakened to "either $D^+f(x) \geqslant 0$ or $D^-f(x) \geqslant 0$". Indeed there exists a counterexample of a continuous function f on $[a, b]$, which is not monotonic, but has the property that for each $x \in (a, b)$ either $D^+f(x) = +\infty$ or $D^-f(x) = +\infty$. Such an example can be found in § 274, 275 of [**4**], but it is too involved to present here.

References

1. G. Aumann, Reelle Funktionen, Die Grundlagen der mathematischen Wissenschaften, Vol. 68, Springer-Verlag, Berlin-Göttingen-Heidelberg, 1954.

2. T. M. Flett, Some historical notes and speculations concerning the Mean Value Theorems of differential calculus, Bull. Inst. Math. Appl., vol. 10 (1974) 66–72.

3. B. R. Gelbaum and M. H. Olmsted, Counterexamples in Analysis, Holden-Day, San Francisco, 1964.

4. E. W. Hobson, The Theory of Functions of a Real Variable, Vol. 2, Dover, New York, 1957.

5. J. W. Knight, Functions with zero right derivatives are constant, this MONTHLY, vol. 87, pp. 657–658.

6. L. Scheefer, Zur Theorie der Stetigen Funktionen einer Reellen Veränderlichen, Acta Math., 5 (1884) 183–194, 279–296.

7. B. Sz.-Nagy, Introduction to Real Functions and Orthogonal Polynomials, Akademiai Kiado, Budapest, 1964.

8. C. G. Young, A note on derivatives and differential coefficients, Acta Math., vol. 37 (1914) 144.

AN INVOLUTION OF BLOCKS IN THE PARTITIONS OF n

A. H. M. HOARE

Department of Mathematics, University of Birmingham, Birmingham B15 2TT, England

Let n be any natural number, then any partition of n can be represented by $(n_1^{a_1}, n_2^{a_2}, \dots, n_r^{a_r})$, where

$$a_1n_1 + a_2n_2 + \cdots a_rn_r = n.$$

If we impose the condition $n_1 < n_2 < \cdots n_r$, then this representation is unique and it is shown in [1] that

$$(*) \qquad \prod a_1!a_2! \cdots a_r! = \prod n_1^{a_1}n_2^{a_2} \cdots n_r^{a_r},$$

where both products are taken over all partitions of n. We show that *there is a natural bijection between the factors of these two products.*

Let π be the partition of n given by

$$(n_1^{a_1}, n_2^{a_2}, \dots, n_r^{a_r}) \quad \text{where} \quad n_1 < n_2 < \cdots < n_r.$$

We say that π has a_1 blocks with n_1 elements, a_2 blocks with n_2 elements, etc. Take any block in π, the kth block with l elements, say, so that $l = n_i$ and $1 \leqslant k \leqslant a_i$ for some i, $1 \leqslant i \leqslant r$. Then in the left-hand side of $(*)$ k is a factor of $a_i!$ contributed by π. Given π, k and $l = n_i$ as above, let $\bar{\pi}$ be the partition of n represented by

$$(n_1^{a_1}, n_2^{a_2}, \dots, n_i^{a_i - k}, \dots, n_r^{a_r}, k^l)$$

or by $(m_1^{b_1}, m_2^{b_2}, \dots, m_s^{b_s})$ with $m_1 < m_2 < \cdots < m_s$. Then $k = m_j$ and $1 \leqslant l \leqslant b_j$ for some j, $1 \leqslant j \leqslant s$. Define the map ϕ taking the factor k of $a_i!$ contributed by π to the left-hand side of

($*$) to the lth factor of $m_j^{b_j}$ contributed by $\bar{\pi}$ to the right-hand side. In fact, what we have is a map taking the kth block with l elements in π to the lth block with k elements in $\bar{\pi}$, where $\bar{\pi}$ depends on π, k and l. For example, if π is $(1^3, 2^5, 3^2, 4^1)$, then the 4th block with 2 elements in π is mapped to the 2nd block with 4 elements in $(1^3, 2^1, 3^2, 4^3)$, the 3rd block with 1 element in π is mapped to the 1st block with 3 elements in $(2^5, 3^3, 4^1)$, etc. It is easily seen from the construction that this map is an involution which fixes precisely those blocks with $k = l$, e.g., the 1st block with 1 element and 2nd block with 2 elements but no other block in $(1^3, 2^5, 3^2, 4^1)$.

Returning to ($*$) we see that the factors of the two sides are in one–one correspondence with the blocks in the partitions of n and that the map ϕ corresponds to the involution above. Thus ϕ is a bijection.

Reference

1. M. S. Kirdar and T. H. R. Skyrme, On an identity relating to partitions and repetitions of parts, Canad. J. Math., 34 (1982) 194–5.

THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

For instructions about submitting material for publication in this department see the inside front cover.

CONVERGENCE WITH PICTURES

P. J. RIPPON
Faculty of Mathematics, The Open University, Milton Keynes, United Kingdom

Two pictures are presented, each of which serves to demonstrate the convergence of an infinite series. No claim to originality is made but the author is surprised not to have found these in any standard analysis text book. Fig. 1 needs no explanation.
It shows that

$$\sum_{n=1}^{\infty} \frac{1}{n^2} < 2.$$

For an elementary proof that this sum is $\pi^2/6$ we refer to [**3**].

Fig. 2 concerns the technique of estimating sums by using integrals. Suppose that the function

$$f\colon [1, \infty) \to (0, \infty)$$

is either convex decreasing or concave increasing and let

$$a_n = \left| \frac{1}{2}(f(n) + f(n+1)) - \int_n^{n+1} f(x)\, dx \right|, \qquad n = 1, 2, \ldots,$$

be the area of the set A_n lying between $y = f(x)$ and the chord joining $(n, f(n))$ to $(n + 1, f(n + 1))$.
The assumptions about f allow us to translate the sets A_n to the nonoverlapping positions indicated in the figure, and this shows that

$$\sum_{n=1}^{\infty} a_n < \frac{1}{2} |f(1) - f(2)|.$$

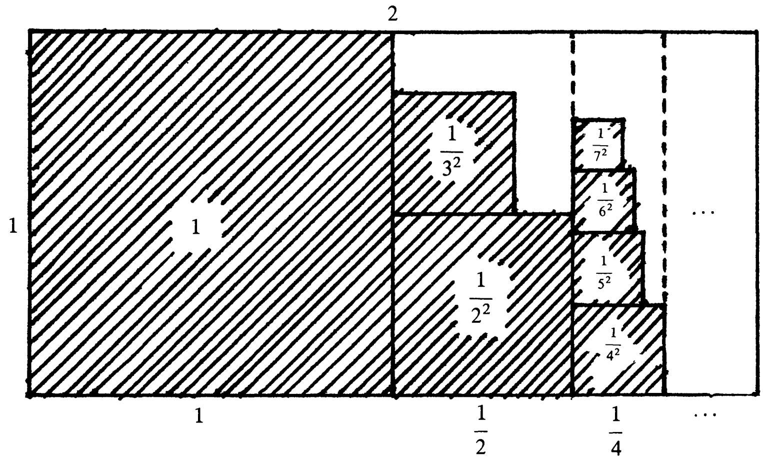


FIG. 1

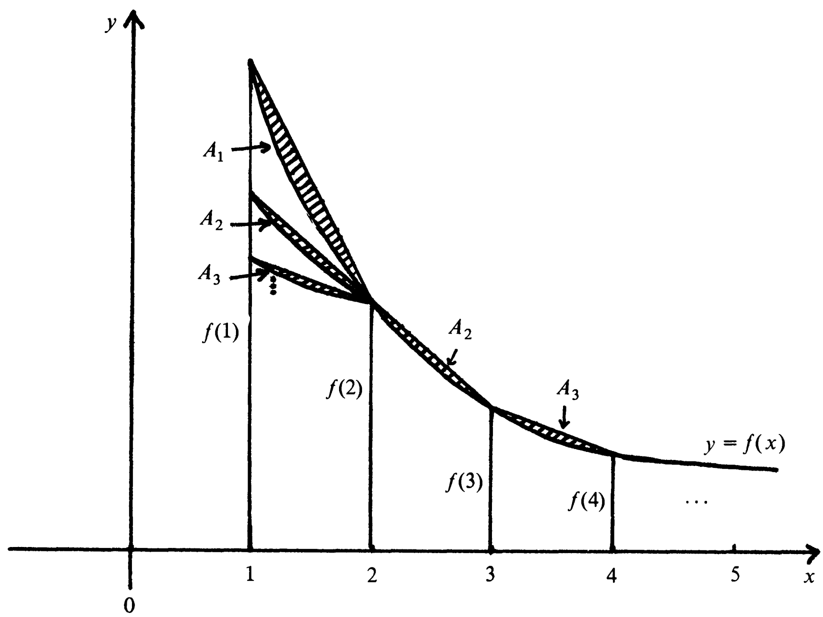


FIG. 2

Now, for $n = 2, 3, \ldots,$ we have

$$(1) \qquad \sum_{k=1}^{n-1} a_k = \left| \sum_{k=1}^{n} f(k) - \frac{1}{2}(f(1) + f(n)) - \int_1^n f(x)\, dx \right|,$$

and so, by taking particular functions f in (1), we obtain various familiar limits. For example, if f is convex decreasing with $\lim_{x \to \infty} f(x) = 0$, and

$$(2) \qquad d_n = \sum_{k=1}^{n} f(k) - \int_1^n f(x)\, dx, \qquad n = 1, 2, \ldots,$$

then $\gamma_f = \lim_{n \to \infty} d_n$ exists. The special case $f(x) = 1/x$ corresponds, of course, to the existence of Euler's constant γ.

The assumption here that f is convex is not essential to the proof that $\lim_{n\to\infty} d_n$ exists. This result holds whenever f is positive and decreasing. Our method, however, allows us to take $f(x) = \log x$ in (1) and we obtain, after a short calculation, the existence of the limit in Stirling's formula

$$\alpha = \lim_{n\to\infty}\left(\frac{n!}{n^{n+1/2}e^{-n}}\right).$$

The proof that $\alpha = \sqrt{2\pi}$ uses Wallis' product (see, for example, [2, p. 363]).

This method of rearranging areas also gives good estimates for the speed of convergence in these limits. For example, if f is convex decreasing with $\lim_{x\to\infty} f(x) = 0$, we obtain rather easily that

$$\text{(3)}\qquad \frac{1}{2}f(n+1) < d_n - \gamma_f < \frac{1}{2}f(n), \qquad n = 1, 2, \ldots .$$

Indeed, by considering only the sets $A_n, A_{n+1}, \ldots$, in Fig. 2 we see that

$$\text{(4)}\qquad \sum_{k=n}^{\infty} a_k < \frac{1}{2}|f(n) - f(n+1)|, \qquad n = 1, 2, \ldots .$$

Now, for $N = n+1, n+2, \ldots$, we have

$$\sum_{k=n}^{N-1} a_k = \left| \sum_{k=n+1}^{N} f(k) + \frac{1}{2}(f(n) - f(N)) - \int_n^N f(x)\,dx \right|,$$

and so, if f is convex decreasing with $\lim_{x\to\infty} f(x) = 0$, then

$$0 < \gamma_f - d_n + \frac{1}{2}f(n) < \frac{1}{2}(f(n) - f(n+1)),$$

since

$$\gamma_f - d_n = \lim_{N\to\infty}\left(\sum_{k=n+1}^{N} f(k) - \int_n^N f(x)\,dx \right).$$

This proves (3).

In fact the lower estimate in (3) can be improved (see [1]) to $\frac{1}{2}f(n+\frac{1}{2})$. This improvement can also be demonstrated using the method of rearranging areas. We leave it as an amusing exercise for the reader to find a way of dissecting the sets $A_n, A_{n+1}, \ldots$, and rearranging the pieces so that they all lie, without overlapping, in a rectangle with sides $\frac{1}{2}$ and $|f(n) - f(n+\frac{1}{2})|$. This shows that the upper estimate in (4) can be improved to $\frac{1}{2}|f(n) - f(n+\frac{1}{2})|$, which gives the desired improvement of the lower estimate in (3).

References

1. R. P. Boas, Estimating remainders, Math. Mag., 51 (1978) 83–89.
2. R. Courant, Differential and Integral Calculus, Vol. 1, Blackie and Son, 1966.
3. I. Papadimitriou, A simple proof of the formula $\sum_{k=1}^{\infty} k^{-2} = \pi^2/6$, this Monthly, 80 (1973) 424–425.

ANSWER TO PHOTO ON PAGE 456

Lipót Fejér (1880–1959).

A MNEMONIC FOR AREAS OF POLYGONS

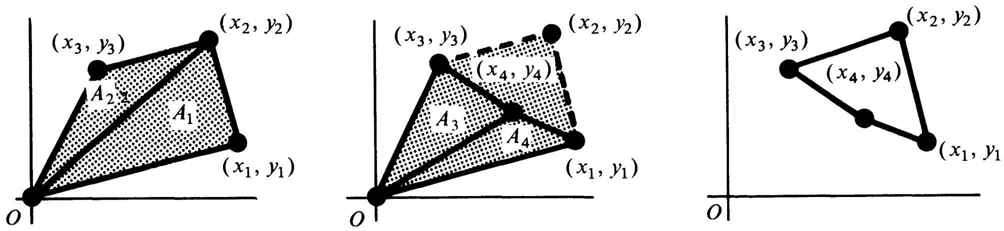
M. G. STONE
Faculty of Science, University of Calgary, Calgary, Alberta, Canada T2N 1N4

Many occasions arise which require an area calculation for some polygonal region with given vertices. Interpretation of 2×2 determinants as areas leads easily to a very useful formula for areas of arbitrary polygonal regions in the plane. This formula, which enables one to write down by inspection an expression for the area, does not seem to be known to most undergraduates. Our note is intended to bring the method to the attention of a wider audience (though no doubt many readers are already familiar with it (cf. [3])), and to provide geometric insight into the mechanism at work.

Recall that

$$\frac{1}{2}\det\begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix} = \frac{1}{2}(x_1y_2 - x_2y_1)$$

is the area of the triangle OBD in Fig. 2(a) if OBD describes a counterclockwise path, and otherwise is the negative of that area. This triangle is the region subtended by $\overline{BD}$ viewed from O.



1(a) Exterior polygon 1(b) Interior polygon 1(c) Required polygon

FIG. 1

The difference in areas subtended by the "exterior" and "interior" boundaries for the polygon in Fig. 1 viewed from the origin O gives the polygonal area

$$A = (A_1 + A_2) - (A_3 + A_4) = \frac{1}{2}\begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} + \frac{1}{2}\begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix} + \frac{1}{2}\begin{vmatrix} x_3 & x_4 \\ y_3 & y_4 \end{vmatrix} + \frac{1}{2}\begin{vmatrix} x_4 & x_1 \\ y_4 & y_1 \end{vmatrix}$$

expressed in terms of triangular regions subtended by edges of the boundary. This can be conveniently written

$$A = \frac{1}{2}\begin{vmatrix} x_1 & x_2 & x_3 & x_4 & x_1 \\ y_1 & y_2 & y_3 & y_4 & y_1 \end{vmatrix}$$

if we interpret $\begin{vmatrix} x_1 & x_2 & x_3 & x_4 & x_1 \\ y_1 & y_2 & y_3 & y_4 & y_1 \end{vmatrix}$ by summing the products on the "downward" diagonals and subtracting the products on the "upward" diagonals. More generally for any counterclockwise enumeration of the vertices of a polygonal region $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, the area is obtained by summing over determinants which express that area as the sum and difference of triangular regions subtended by the boundary viewed from the origin. The area is given by

$$A = \frac{1}{2}\begin{vmatrix} x_1 & x_2 & \cdots & x_n & x_1 \\ y_1 & y_2 & \cdots & y_n & y_1 \end{vmatrix},$$

where the determinant like symbol (an economical way of writing a sum of 2×2 determinants) is interpreted in the manner indicated above. A little investigation will reveal that it is *not* necessary for the polygon to be convex; the formula conveniently works for any oriented polygonal region

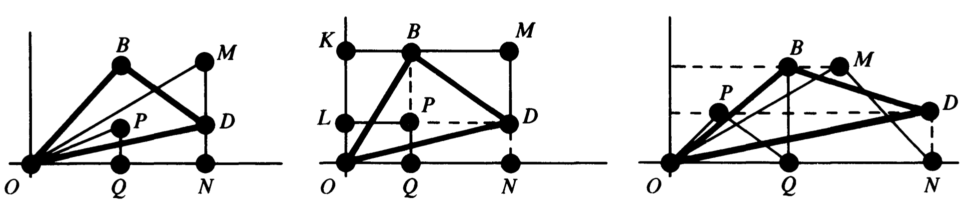
in the plane without regard to (quadrant) location of the vertices or convexity.

The observation that $\frac{1}{2}\begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix}$ is the (signed) area of the triangle OBD in Fig. 2 is often made by decomposing the region $OBDE$ (in Fig. 2(b)) to obtain

$$OBD = \frac{1}{2}x_1y_1 + \frac{1}{2}(y_1 + y_2)(x_2 - x_1) - \frac{1}{2}x_2y_2 = \frac{1}{2}x_2y_1 - \frac{1}{2}x_1y_2 = -\frac{1}{2}\begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix}.$$

2(a) 2(b)

FIG. 2



3(a) $(OMN - OPQ) = OBD$ 3(b) $\frac{1}{2}(OKMN - OLPQ) = OBD$ 3(c) $(OMN - OPQ) = OBD$

FIG. 3

An interesting *geometric* fact however underlies the algebraic relationship $OBD = \frac{1}{2}x_2y_1 - \frac{1}{2}x_1y_2$, as can be seen in Fig. 3. The relation in Fig. 3(a) is a special case of the more general relationship between areas of triangles indicated in Fig. 3(c), $OBD = OMN - OPQ$. Note that 3(a) follows easily from 3(b) which is itself an interesting observation about rectangles! Both 3(b) and 3(c) are easy consequences of the fact that triangles with equal bases and equal altitudes have equal areas; a synthetic geometric proof of each of the above assertions is a nice student exercise. Calculus instructors will find the formula itself useful in calculations involving the trapezoidal rule. At a more advanced level the derivation of this formula from its analytic counterpart (Green's Theorem for line integrals) can be explored (cf. [**2**]). Computer science graphics texts which utilize trapezoidal decomposition (cf. [**1**]) could benefit from incorporation of the formula in discussing areas of planar regions.

References

1. Foley and VanDam, Fundamentals of Computer Graphics, p. 515, Addison-Wesley, 1983.
2. Marsden and Tromba, Vector Calculus, p. 409, Freeman, 1976.
3. Underwood and Sparks, Analytic Geometry, 3rd ed., Houghton Mifflin, Boston, 1963.

COUNTABILITY OF SETS

STEPHEN L. CAMPBELL

Department of Mathematics, North Carolina State University, Raleigh, NC 27650

In an introductory course, the countability of the rational numbers is usually proved using a diagonalization argument. There is an alternative approach that is, perhaps, more intuitive. The idea is not mine; I heard it in graduate school 15 years ago. However, since then I have never met

anyone else who has heard of it.

We shall say that two sets have the same cardinality if each can be mapped in a one-to-one manner into the other. A set is countable if it has the same cardinality as the integers.

THEOREM 1. *The set of rational numbers is countable.*

Proof. Clearly the integers can be mapped into the rationals. Now observe that each rational number a/b is a distinct integer written in the base 11 with / as the extra symbol for 10. Q.E.D.

For example,

$$2/3 = 2(11^2) + 10(11^1) + 3(11^0) = 355.$$

Note that we have actually shown that the set of all representations of rational numbers is countable.

By writing expressions in "word processor notation" and using larger bases one can trivially show many other sets are countable. For example, the set of polynomials over the rationals can be mapped into the integers using base 16 with symbols

$$\{0,1,2,3,4,5,6,7\,8,9,/,+,-,\%,\&,x\},$$

where % means start superscript, & means end superscript, and x is the independent variable.

PROBLEMS AND SOLUTIONS

EDITED BY G. L. ALEXANDERSON, H. M. W. EDGAR (ELEMENTARY PROBLEMS), DALE H. MUGLER, AND KENNETH B. STOLARSKY (ADVANCED PROBLEMS)

*For instructions about submitting **proposed** problems for publication in this department see the inside front cover. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the address given on the inside front cover.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

For instructions about submitting solutions of these Elementary Problems, which should be mailed by November 30, 1986, see the inside front cover. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgement).

E 3153. *Proposed by Lászlo Cseh and Imre Merényi (students), Babes-Bolyai University, Cluj, Romania.*

For integers $n \geqslant 2$, $r \geqslant 0$, let $S_n(r) = \sum_{k=1}^{\infty} k^r n^{-k}$. Show that for all $r \geqslant 1$,

$$S_n(r) = \frac{n}{n-1}\left[\binom{r}{1}S_n(r-1) - \binom{r}{2}S_n(r-2) + \cdots + (-1)^{r+1}\binom{r}{r}S_n(0)\right].$$

E 3154. *Proposed by George A. Tsintsifas, Thessaloniki, Greece.*

Let A_1, B_1, C_1 be points on the sides a, b, c of the triangle ABC, respectively, and a_1, b_1, c_1 the sides of the triangle $A_1B_1C_1$. Prove

$$a^2b_1c_1 + b^2c_1a_1 + c^2a_1b_1 \geqslant 4F^2,$$

where F is the area of triangle ABC.

E 3155. *Proposed by Gene Bennett, John Glenn, and Clark Kimberling, University of Evansville.*

Prove that for any triangle ABC, there exist points A', B', C' satisfying

(1) A' lies on side $\overline{BC}$, B' on side $\overline{AC}$, and C' on side $\overline{AB}$;
(2) $\underline{A'C} + \underline{CB'} = \underline{B'A} + AC' = C'B + BA'$; and
(3) $\overline{AA'}$, $\overline{BB'}$, and $\overline{CC'}$ concur in a point.

E 3156. *Proposed by Raphael M. Robinson, University of California, Berkeley.*

Suppose that r, s, t are integers with $r \geqslant 0$, $s \geqslant 0$, $t = r + s \geqslant 2$. Is there a word W of length t in the alphabet $\{a, b\}$ such that $W = AB = Cab$, where A, B, C are palindromes, and the lengths of A and B are r and s? Show that such a word W exists if and only if $r + 2$ is prime to $s - 2$, and that in this case it is unique.

E 3157. *Proposed by Liviu I. Nicolaescu, University Al. I. Cuza, Iassy, Romania.*

How many sets of four distinct points forming the vertices of a trapezoid are there if the points are chosen from the vertices of a regular n-gon ($n \geqslant 4$)?

E 3158. *Proposed by I. J. Schoenberg, Madison, WI.*

Consider the image C_n of the circle $z = e^{it}$ $(0 \leqslant t \leqslant 2\pi)$ under the function $f(z) = z + az^n$, where a is a real constant. Show that
(a) if $|a| > 0$, then C_n is a closed Jordan curve if and only if $|a| \leqslant 1/n$,
(b) if $n \geqslant 2$, then C_n is a closed convex curve if and only if $|a| \leqslant 1/n^2$.

SOLUTIONS OF ELEMENTARY PROBLEMS

Evaluation of a lim sup

E 2921 [1982, 63]. *Proposed by P. R. Halmos, University of Santa Clara.*

Evaluate

$$\limsup_{n\to\infty} |(2+3i)^n - (2-3i)^n|^{1/n}.$$

The problem looks frightening in this form. To make it look less so, put

$$r(a, b) = \limsup_{n\to\infty} |a^n - b^n|^{1/n}$$

for each pair of complex numbers a, b. If $a = b$, then $r(a, b) = 0$; what is it if $a \neq b$?

Solution I by G. A. Heuer, Concordia College, Moorhead, MN. If $a \neq b$, $r(a, b) = \max\{|a|, |b|\}$. This is trivial if a or b is zero, so suppose that both are nonzero and, without loss of generality, that $|a| \geqslant |b|$. Write

$$r(a, b) = |a| \limsup_{n\to\infty} |1 - (b/a)^n|^{1/n}.$$

If $|b| < |a|$, the result is now obvious. If $|b| = |a|$ we may write $b/a = e^{i\theta}$, and then

$$2 \geqslant |1 - (b/a)^n| = |1 - e^{in\theta}| = (2 - 2\cos(n\theta))^{1/2} \geqslant 1$$

whenever $e^{in\theta}$ does not fall in the sector of the complex plane between $-\pi/3$ and $\pi/3$. Since this condition includes more than half of the unit circle, it will hold for infinitely many n unless $\theta = 0$; i.e., unless $b = a$. The claim follows.

Solution II *by Timothy S. Norfolk, The University of Akron.* For a and b nonzero and $a \neq b$, consider the function

$$f(z) = \frac{1}{1 - az} - \frac{1}{1 - bz},$$

which is analytic for

$$|z| < \frac{1}{\max\{|a|, |b|\}} = R,$$

and has the power series expansion

$$f(z) = \sum_{n=0}^{\infty} (a^n - b^n) z^n$$

in this region. By the Cauchy-Hadamard Theorem,

$$\limsup_{n \to \infty} |a^n - b^n|^{1/n} = \frac{1}{R} = \max\{|a|, |b|\}.$$

Also solved by fifty-four other readers and the proposer.

An Application of Newton's Formulae

E 2993 [1983, 287]. *Proposed by Michael Larsen, Student, Harvard University.*

Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be complex numbers such that $\Sigma_1^n \alpha_i^m$ is an integer for every positive integer m. Show that the polynomial $\Pi_1^n (x - \alpha_i)$ has integer coefficients.

Solution I *by A. A. Jagers, Technische Hogeschool Twente, The Netherlands.* Let a_i be the coefficient of x^{n-i} in $\Pi_1^n(x - \alpha_i)$. Then $g(x) = \Pi_1^n(1 - \alpha_i x)$ is equal to $a_0 + a_1 x + \cdots + a_n x^n$ with $a_0 = 1$. Let $h(x)$ be a generating function of the power sums $s_k = \Sigma_{i=1}^n \alpha_i^k$; $h(x) = \Sigma_{k=1}^{\infty} s_k x^{k-1}$, say. Then

$$h(x) = -(d/dx)\log g(x), \quad \text{or} \quad g'(x) + h(x) g(x) = 0,$$

or, by equating coefficients of equal powers of x,

$$m a_m + s_1 a_{m-1} + s_2 a_{m-2} + \cdots + s_m a_0 = 0 \quad (1 \leqslant m \leqslant n).$$

Since $s_k \in \mathbb{Z}$, it follows by induction on m that $m! a_m \in \mathbb{Z}$ for all m. Then surely $n! a_m \in \mathbb{Z}$ and so $n! \alpha_i$ is an algebraic integer for all i. However, since $s_k \in \mathbb{Z}$ for *all* k, the above argument may be repeated with α_i^k instead of α_i. It follows that $n! \alpha_i^k$ is an algebraic integer for all k; that is, α_i is an algebraic integer itself. Hence so are the coefficients a_m being integer polynomials in the α_i. Since also $a_m \in \mathbb{Q}$, we have $a_m \in \mathbb{Z}$ for all m.

Solution II *by Ira Gessel, Massachusetts Institute of Technology.* We shall show that the coefficients are algebraic integers and rational numbers.

To allow for repetition among the α_i, we show first that for any nonzero complex numbers $c_1, \ldots, c_n$ and distinct complex numbers $\beta_1, \ldots, \beta_n$, if $\Sigma_1^n c_i \beta_i^m$ is an integer for every positive integer m, then each β_i is an algebraic integer. Let V be the Vandermonde matrix

$(c_i\beta_i^{j-1})_{i,j=1,\ldots,n}$. Then for any positive integer k, the $1 \times n$ matrix $(\beta_1^k\beta_2^k \cdots \beta_n^k)V$ has integer entries. Multiplying on the right by V^{-1}, which must exist, we find that β_i^k is a linear combination with integer coefficients of the entries of V^{-1}. Thus $\mathbb{Z}[\beta_i]$ is a finitely generated $\mathbb{Z}$-module, so β_i is integral over $\mathbb{Z}$.

It follows that the α_i are algebraic integers, hence so are the coefficients of $\Pi(x - \alpha_i)$.

Now let $\gamma_1, \ldots, \gamma_n$ be arbitrary complex numbers, let

$$\sum_1^n \gamma_i^m = p_m,$$

and let

$$\prod_1^n (x - \gamma_i) = x^n + \sum_1^n (-1)^i e_i x^{n-i}.$$

It is well known that each e_i is a polynomial in the p_m with rational coefficients. This is easily seen by equating coefficients of x^i in

$$1 + \sum_1^n e_i x^i = \exp\log\prod_1^n (1 + \gamma_i x) = \exp\left(\sum_1^\infty \frac{p_i}{i}(-1)^{i-1}x^i\right).$$

It follows that the coefficients of $\Pi(x - \alpha_i)$ are rational numbers, and are therefore integers.

Also solved by M. Bencze (Romania), D. Richman, University of South Alabama Problem Group, and the proposer.

Approximating a Sine Function on a Square

E 3006 [1983, 401]. *Proposed by J. R. Kuttler, Johns Hopkins University.*

Find a function of the form $\alpha(x) + \beta(y)$ which best approximates $\sin x \sin y$ on the square $S = \{(x, y)\colon 0 \leqslant x \leqslant 2\pi, 0 \leqslant y \leqslant 2\pi\}$ in the sup norm, i.e., so that

$$\sup_{(x,y)\in S} |\sin x \sin y - \alpha(x) - \beta(y)|$$

is as small as possible.

Solution by Raymond E. Rogers, Oakland, CA. Answer: $\alpha(x) = \beta(y) = 0$. Suppose a better approximation, $a(x)$, $b(y)$, exists. Then

$$a\left(\frac{\pi}{2}\right) + b\left(\frac{\pi}{2}\right) > 0, \tag{1}$$

$$a\left(\frac{3\pi}{2}\right) + b\left(\frac{\pi}{2}\right) < 0, \tag{2}$$

$$a\left(\frac{3\pi}{2}\right) + b\left(\frac{3\pi}{2}\right) > 0, \tag{3}$$

$$a\left(\frac{\pi}{2}\right) + b\left(\frac{3\pi}{2}\right) < 0. \tag{4}$$

(1)–(2) yields $a\left(\frac{\pi}{2}\right) - a\left(\frac{3\pi}{2}\right) > 0$, (4)–(3) yields $a\left(\frac{\pi}{2}\right) - a\left(\frac{3\pi}{2}\right) < 0$, a contradiction.

Also solved by U. Abel (West Germany), P. G. de Buda (Canada), W. Castrellon (Colombia), S. F. Henderson, G. A. Heuer, K. Kearnes, B. G. Klein, D. Lindsay, O. P. Lossers (The Netherlands), V. D. Mascioni (Switzerland), C. A. Meyer (West Germany), J.-M. Monier (France), S. Noltie, I. Pokorny (Czechoslovakia), J. Propp, T. J. Rivlin, A. J. Schwenk, R. Stong, University of South Alabama Problem Group, M. Wilson, P. Y. Wu (Taiwan), and the proposer.

ADVANCED PROBLEMS

For instructions about submitting solutions of these Advanced Problems, which should be mailed by November 30, 1986, see the inside front cover. The solver's full post-office address should be on each sheet.

6521. *Proposed by C. S. Gardner and C. Radin, University of Texas at Austin.*

Let a, b be fixed, with $0 < a < b$, and let f be a variable function, subject to

$$f(x) \geqslant 0, \quad -\infty < x < \infty, \quad \text{and} \quad \int_{-\infty}^{\infty} f(x)\,dx = 1.$$

What is the maximum value of

$$V(f) = \int_{-\infty}^{\infty}\int_{y+a}^{y+b} f(x)f(y)\,dx\,dy?$$

6522. *Proposed by Gunnar Blom, University of Lund and Lund Institute of Technology, Sweden.*

Let $X_1, X_2, \ldots$ be an infinite sequence of independent random variables with the common continuous distribution function F. Let X_N be the first variable that is less than exactly one of all its predecessors $X_1, \ldots, X_{N-1}$. Determine the distribution function of X_N.

6523. *Proposed by David G. Cantor, University of California, Los Angeles.*

Suppose $f(z)$ is an irreducible polynomial of degree d over the field of rational numbers, and suppose that $f(z)$ has two roots α, β with α/β a primitive nth root of unity. Show that $\phi(n) \leq d$.

SOLUTIONS OF ADVANCED PROBLEMS

A Digamma Determinant

6474 [1984, 588]. *Proposed by Heinz-Jürgen Seiffert, Berlin, West Germany.*

Let $\Psi_n(z)$ be the $n \times n$ matrix $(\psi(z+j+k))_{j,k=0,1,\ldots,n}$, where $\psi(z) = \Gamma'(z)/\Gamma(z)$ is the digamma function. Prove that

$$\det \Psi_n(z) = \left(\prod_{j=0}^{n-1}(j!)^2\right)\sum_{k=0}^{n}(-1)^k\binom{n}{k}\psi(z+k)\prod_{\substack{r=0\\ r\neq k}}^{n}\prod_{j=0}^{n-1}(z+j+r)^{-1}.$$

Solution by James C. Smith, University of South Alabama, Mobile, AL. The digamma function satisfies

$$\psi(z+1) - \psi(z) = z^{-1}.$$

Set $a_{jk} = (z+j+k-1)^{-1}$, $j = 1,2,\ldots,n$, and set $a_{0k} = \psi(z+k)$. Then $A = (a_{jk})$ is the matrix formed by subtracting the $(j-1)$st row from the jth row of $\Psi_n(z)$ sequentially for $j = n,\ldots,2,1$. Thus,

$$\det \Psi_n(z) = \det A = \sum_{k=0}^{n}(-1)^k\psi(z+k)\det A_k,$$

where A_k is the cofactor of A given by (a_{jr}), where $j = 1,2,\ldots,n$ and $r = 0,1,\ldots,n$, $r \neq k$.

Observe that A_k is a case of the double alternant

$$\Delta = \left|(\alpha_j - \beta_u)^{-1}\right|,$$

where $j, u = 1,2,\ldots,n$, the α's have the values $n+1, n+2,\ldots,2n$, and the β's the values

$$n+1-z,\, n-z,\, \ldots,\, n-k+2-z,\, n-k-z,\, n-k-1-z,\, \ldots, 1-z.$$

The formula

$$\Delta = (-1)^{n(n-1)/2} P^{1/2}(\alpha_1, \ldots, \alpha_n) P^{1/2}(\beta_1, \ldots, \beta_n) \prod_{j=1}^{n} \prod_{u=1}^{n} (\alpha_j - \beta_u)^{-1},$$

where

$$P^{1/2}(x_1, \ldots, x_n) = \prod_{i>j} (x_i - x_j),$$

is well known.

In the case of A_k we have

$$P^{1/2}(\alpha_1, \ldots, \alpha_n) = \prod_{j=1}^{n-1} j!$$

and

$$P^{1/2}(\beta_1, \ldots, \beta_n) = (-1)^{n(n-1)/2} \left(\prod_{j=1}^{n-1} j! \right) \binom{n}{k}.$$

Since

$$\prod_{j=1}^{n} \prod_{u=1}^{n} (\alpha_j - \beta_u) = \prod_{j=0}^{n-1} \prod_{\substack{r=0 \\ r \neq k}}^{n} (z + j + r),$$

the result follows.

By similar techniques Robert E. Shafer has evaluated the n-variable expression $\det(a_{ij})$, where $a_{ij} = \psi(z_i + i + j - 2)$, $1 \le i, j \le n$.

Also solved by Syrous Marivani and the proposer.

Polynomial Bounds for Diophantine Equations

6479. *Proposed by D. W. Masser, The University of Michigan, Ann Arbor.*

Let a, b, c, d be integers such that the polynomial $P(x) = x^4 + ax^3 + bx^2 + cx + d$ is not a perfect square. A well-known theorem of Siegel implies that the diophantine equation $y^2 = P(x)$ has only a finite number of solutions in integers x, y. For $H \geqslant 1$ let $X(H)$ be the maximum of $|x|$ taken over all such solutions of all such equations with $\max(|a|, |b|, |c|, |d|) \leqslant H$. Show that there are absolute constants $k > 0$, K such that $kH^3 \leqslant X(H) \leqslant KH^3$ for all H.

Solution by the proposer. Though not best possible, the constants $k = 2^{-12}$ and $K = 26$ suffice.

We begin with the upper bound. Put $e = 4b - a^2$ and

$$C = 64c - 8ae, \qquad D = 64d - e^2.$$

Since $\max(|a|, |b|, |c|, |d|, 1) \leqslant H$, we have $|e| \leqslant 5H^2$. Put also

$$Q(x) = 8x^2 + 4ax + e;$$

then we shall use the identity

$$(*) \qquad 64P(x) - (Q(x))^2 = Cx + D.$$

Let x be an integer such that $P(x) = y^2$ for some integer y. To estimate $|x|$ we split into five cases.

(I) $|x| \leqslant H$. Trivial.

(II) $|x| \geqslant H$, $Cx + D \neq 0$. Now we use the easy inequality

$$|u^2 - v^2| \geqslant |v|$$

valid for any integers u, v with $u^2 \neq v^2$. For $u = 8y$, $v = Q(x)$ we deduce from (*)

$$|Q(x)| \leqslant |Cx + D|.$$

But $|C| \leqslant 104H^3$, $|D| \leqslant 89H^4$, so we get

$$8|x|^2 \leqslant 4H|x| + 5H^2 + 104H^3|x| + 89H^4 \leqslant 202H^3|x|,$$

whence

$$|x| \leqslant 26H^3.$$

(III) $\boldsymbol{Cx + D = 0, a = 0}$. Now $C \neq 0$, else $D = 0$ as well and the polynomial P would be a square from (*). So $|C| \geqslant 8$, and since now

$$|D| = |64d - 16b^2| \leqslant 80H^2$$

we get

$$|x| \leqslant 10H^2.$$

(IV) $\boldsymbol{Cx + D = 0, a \neq 0, |e| \geqslant 12H}$. Now

$$|C| \geqslant |8ae| - |64c| \geqslant 8|e| - 64H \geqslant |e|$$

and

$$|D| \leqslant 64H + |e|^2 \leqslant 2|e|^2.$$

This yields

$$|x| = |D/C| \leqslant 2|e|$$

or

$$|x| \leqslant 10H^2.$$

(V) $\boldsymbol{Cx + D = 0, |e| \leqslant 12H}$. As above $C \neq 0$, so $|C| \geqslant 8$, and now

$$|D| \leqslant 64H + |e|^2 \leqslant 208H^2.$$

This yields

$$|x| \leqslant 26H^2.$$

Upon combining all five cases, we get

$$X(H) \leqslant 26H^3$$

for all H.

REMARK. The method used is essentially that of Runge (J. Reine Angew. Math., 100 (1887), 425–435), simplified out of recognition in a special case. For related results see a paper of D. L. Hilliker (Math. Comp., 38 (1982), 611–626), and the references therein to forthcoming work of the author and E. G. Straus, Trans. Amer. Math. Soc., 280 (1983), 637–657).

We now obtain the lower bound by considering two cases.

(I) $\boldsymbol{1 \leqslant H \leqslant 16}$. Since $P(1) = 0^2$ for $P(x) = x^4 - 1$, it is obvious that $X(H) \geqslant X(1) \geqslant 1$. Hence

$$X(H) \geqslant H^3/4096.$$

(II) $\boldsymbol{H \geqslant 16}$. Write $t = [H/8] \geqslant H/16$ and

$$P_t(x) = x^4 + 8tx^3 - 12x^2 + 4.$$

Since $P_t(-1) + P_t(1) = -14 < 0$, it is clear that $P_t(x)$ is not a perfect square. But it can be

verified that

$$P_t(4t^3 - 2t) = (16t^6 - 12t^2 + 2)^2.$$

Also $t \geqslant 2$, and so

$$X(H) \geqslant X(8t) \geqslant 4t^3 - 2t \geqslant 2t^3.$$

Thus

$$X(H) \geqslant H^3/2048.$$

Upon combining the two cases, we get

$$X(H) \geqslant H^3/4096$$

for all H.

Resource Sharing for Efficiency

6482* [1984, 652]. *Proposed by G. Vrancken, Utrecht, Holland.*

Let $R(a,p) = \int_0^\infty e^{-ax}(1+x)^p\,dx$. Prove or disprove that

$$\frac{1}{R(a,p)} + \frac{1}{R(b,q)} \geqslant \frac{1}{R(a+b,p+q)}$$

for $a, b > 0$ and $p, q \geqslant 0$.

Partial solution by Donald R. Smith and Ward Whitt, AT & T Bell Laboratories, Holmdel, New Jersey. Assume that p and q are *integers*. Upon expanding $(1+x)^p$ and integrating, we obtain

$$R(a,p) = \sum_{j=0}^{p} \binom{p}{j} j!\,a^{-j-1},$$

from which it is not hard to establish the identity

$$a - \frac{1}{R(a,p)} = \frac{\sum_0^p jA_j}{\sum_0^p A_j}, \quad \text{where } A_j = A_j(a) = a^j/j!.$$

Now set $r = p + q$, $C_j = (a+b)^j/j!$, and

$$P_j = \sum_{n+m=j} \frac{a^n}{n!}\frac{b^m}{m!},$$

where the indices satisfy $0 \leqslant n \leqslant p$ and $0 \leqslant m \leqslant q$, respectively. Clearly

$$a + b - \frac{1}{R(a+b,p+q)} = \frac{\sum_0^r jC_j}{\sum_0^r C_j} =: C.$$

By grouping terms (formed by products of sums) according to the sum of their indices, we obtain

$$a - \frac{1}{R(a,p)} + b - \frac{1}{R(b,q)} = \frac{\sum_0^r jP_j}{\sum_0^r P_j} =: P$$

from the identity. If

$$P_{j+1}/P_j \leqslant C_{j+1}/C_j = \frac{a+b}{j+1},$$

the result $P \leqslant C$ would easily follow by a center of mass consideration. But ("negative factorials are infinite")

$$P_{j+1} = \sum_{n+m=j+1} \binom{n+m}{j+1} \frac{a^n}{n!} \frac{b^m}{m!}$$

$$= \frac{a}{j+1} \sum_{(n-1)+m=j} \frac{a^{n-1}}{(n-1)!} \frac{b^m}{m!} + \frac{b}{j+1} \sum_{n+(m-1)=j} \frac{a^n}{n!} \frac{b^{m-1}}{(m-1)!}.$$

Make the change of variables $N = n - 1$, $M = m - 1$, so that $N \leqslant p - 1 < p$ and $M \leqslant q - 1 < q$; thus

$$P_{j+1} \leqslant \frac{a}{j+1} P_j + \frac{b}{j+1} P_j$$

and the ratio inequality is established.

Since

$$B(p, a) = [aR(a, p)]^{-1}$$

is the Erlang blocking formula in queueing or traffic theory, in physical terms this problem states that the total blocked traffic from separate systems (i.e., when a Erlangs are offered to p trunks in one system and b Erlangs are offered to q trunks in the other) is greater than the blocked traffic from a combined system (when $a + b$ Erlangs are offered to $p + q$ trunks). The problem for general p and q (as well as the convexity of $B(p,a)$ for $p > 0$) remains open; it is widely conjectured by teletraffic engineers and queueing theorists. For further information see D. R. Smith and W. Whitt, *Resource sharing for efficiency in traffic systems*, Bell System Tech. J., 60 (1981), 39–55, and also E. Arthurs and B. W. Stuck, *Problem* 80-19*, SIAM Review 23 (1981), 527–528. A study of $R(a, p)$ as a function of complex variables is given by D. L. Jagerman, *Some properties of the Erlang loss function*, Bell System Tech. J., 53(1974), 525–551, where references to the queueing theory literature are also given.

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER

Actuarial Mathematics. By Newton L. Bowers, Jr., Hans U. Gerber, James C. Hickman, Donald A. Jones, Cecil J. Nesbitt. Society of Actuaries, 500 Park Blvd., Itasca, IL 60143, 1984.

LOUIS M. FRIEDLER
Department of Mathematics, College of St. Scholastica, Duluth, MN 55811

The exams of the Society of Actuaries have undergone fundamental changes in the last few years and the preparation required of actuarial students is now much more mathematically demanding. The importance of the changes, especially in Part 4 of the ten-part Fellowship exams, has been widely publicized to the actuarial community, e.g., [**8**], but not at all, to my knowledge, to the mathematical community.

A major reason mathematics departments should be aware of the changes in the actuarial exams is that actuarial science presently offers abundant opportunities for graduating math majors. Actuarial science can offer an attractive alternative to those bright math students who want to begin work immediately after graduation, communicate well, and are interested in management. (The Society of Actuaries estimates a 40% increase in the number of positions in the next five years [**9**].) If actuarial science is used as a way of attracting students to mathematics, then both those who advise these students and those who teach them specific actuarial courses owe it to the students to be somewhat knowledgeable about the field. Such knowledge should include some idea of the mathematics that is used, an awareness of the current syllabus for the exams, and a realization that actuaries are primarily business people, not applied mathematicians.

A second reason for mathematicians to care about the changes in the exams is that actuarial students now learn more interesting mathematics than they used to. This is especially noticeable in the new book, *Actuarial Mathematics* by Bowers *et al*., that I will discuss in more detail below. The text develops stochastic rather than deterministic models and provides many examples (including examples of sums of random variables, moment generating functions, and Poisson, compound Poisson, and exponential distributions) that are appropriate for an undergraduate course in probability. Moreover, the new book clearly shows fresh interest in mathematics on the part of actuaries, declaring in the introduction, "If actuarial science is to remain in the mainstream of science, it is necessary to recast basic models in the language of contemporary mathematics."

What are the changes in Parts 1–4 of the exams? The first two parts remain the same. The first is based on calculus and linear algebra and has an occasional problem on ordinary differential equations. The second is based on the standard material of an undergraduate course in probability and mathematical statistics. (The one-semester introductory course suggested by the Committee on the Undergraduate Program in Mathematics [**2**] would not, however, be sufficient preparation for this exam.)

The parts of the exams beyond the second are closely tied to specific reading. Part 3 has been updated and improved mathematically. It now includes sections on numerical analysis, operations research, and applied statistics (including regression, analysis of variance, and time series) based on the texts [**1**] (which replaces the mathematically unsatisfactory [**5**]), [**3**], and [**7**].

The most important change in the exams occurs in the content of Part 4. This exam is based on *The Theory of Interest* [**6**] and on Chapters 3–10 and 14–15 of *Actuarial Mathematics* which has replaced *Life Contingencies* by C. W. Jordan, Jr. [**4**], used in two editions by the Society of Actuaries for 30 years. To put that in perspective, imagine all colleges using the same elementary book in any area of mathematics for 30 years.

The text by Jordan is well written but is really concerned with the manipulation of what seems like 10,000 formulas. The model Jordan uses to develop the actuarial functions is based on the "life table." It is assumed that there is a collection of l_0 newborn lives and each of these who reaches age $x-1$ is subject to an effective annual rate of mortality, q_{x-1}. Then we define

$$l_x = l_{x-1}(1 - q_{x-1}) \text{ for } x > 0.$$

The deterministic model of the life table used in Jordan contrasts with the stochastic model built in *Actuarial Mathematics*. Let $T(x)$ represent the continuous random variable: time until death of an individual aged x. After studying the probability density function of this random variable, one considers, again, l_0 newborns but then defines $L(x)$ to be the number of survivors to age x from the l_0 newborns. Then the authors define $l_x = E[L(x)]$, the expected number of survivors to age x of the l_0 newborns. We recover the life table, and so can reproduce the standard actuarial formulas, but we now interpret this table as the average or expected number of lives. The advantage of this interpretation is that we can now calculate variances. For example, to find the premium, P, of an annuity necessary for an insurance, the authors set up a loss random variable as a function of T and P. Setting the expectation of this loss equal to 0 and solving for P yields the classic formula given in Jordan. However, this loss now also has a variance; the larger

the variance, the larger the risk to the company. We could also determine a premium so that the probability of a loss is less than, say, .05. As another application of the advantages of this interpretation, R. McKay mentions in [**8**] that annuities for males and females have long been calculated using separate mortality tables. Now, insurance companies are being pressured to have unisex pricing of annuities. One insurance company has (reportedly) found that when the two mortality tables or distributions are combined, the variance, and so the risk to the company, increases. This information could not have been derived from the techniques in Jordan.

While the approach to contingencies in *Actuarial Mathematics* is substantially different from that in Jordan, many of the topics covered remain the same. The differences include more emphasis on pensions and expenses, integration of the subject matter with risk theory, and the deletion of a chapter on stable populations.

The book under review is well motivated and is suitable for the self-study that is the usual preparation for the Part 4 exam. However, the mathematician reading this book may not find it completely smooth going. Although computation is less stressed than in Jordan, actuaries must be familiar with many standard formulas, and so parts of this book become quite tedious. The authors also regularly misuse the word infinitesimal, as in "infinitesimal time interval of length dt_0". They state in the introduction that the necessary prerequisites include only calculus and undergraduate probability, yet they occasionally use mathematics beyond that level.

Nevertheless, this new approach is a major improvement on Jordan. We can only hope that the questions on the Part 4 exam will eventually change to reflect not only the new material but the new emphasis on understanding.

References

1. R. L. Burden, J. D. Faires, and A. C. Reynolds, Numerical Analysis, 2nd ed., Prindle, Weber and Schmidt, 1981.
2. Committee on the Undergraduate Program in Mathematics, Recommendations for a General Mathematical Sciences Program, Math. Assoc. of Amer., 1981.
3. F. Hillier and G. Lieberman, Introduction to Operations Research, 3rd ed., Holden-Day, 1980.
4. C. W. Jordan, Jr., Life Contingencies, Society of Actuaries, 1952, revised edition, 1967.
5. S. G. Kellison, Fundamentals of Numerical Analysis, Richard. D. Irwin, 1975.
6. ______, The Theory of Interest, Richard D. Irwin, 1970.
7. R. B. Miller and D. W. Wichern, Intermediate Business Statistics, Holt, Rinehart, Winston, 1977.
8. Open Forum: New Textbook on Life Contingencies, Record (Hollywood Meeting), Society of Actuaries, 9 (1983) 2029–2057.
9. W. G. Poortvliet, Survey of demand for actuaries, The Actuary, 18, 3 (1984) 6–7.

Great Moments in Mathematics (after 1650). By Howard Eves. The Mathematical Association of America, 1981 (The Dolciani Mathematical Expositions, Number Seven). xii + 263 pp.

JUDITH V. GRABINER
California State University, Dominguez Hills, Carson, CA 90747

What if a mathematician wants to read about the history of mathematics? Or suppose the Department says, "It's your turn to teach the History of Math course." Going to the bookstore yields the elementary and outdated volumes of D. E. Smith; the undergraduate library supplies the entertaining but untrustworthy *Men of Mathematics* by E. T. Bell; the research library presents shelves of *Histories*, *Histoires*, and *Vorlesungen über Geschichte*. An attempt to find important original papers leads one to Thirteen Books of Euclid's *Elements*, seven volumes of Weierstrass's *Werke*, or an apparently infinite number of volumes of the works of Euler in Latin. Asking a historian for help is apt to provoke a lament that there is nothing suitable to

recommend: no existing collection of sources is adequate, no individual *History* fills one's need. Persistence may convince the historian to give a long list of monographs, each on a different specialized topic and accompanied by a *caveat*, and to recommend articles in various journals in French, German, Italian, and Russian. What, then, is a mathematician to do?

Eves' *Great Moments* is a valiant attempt to provide something that is greatly needed—a combination of narrative history and a collection of specific pieces of past mathematics. The "great moments" are twenty important developments, ranging from "the birth of mathematical probability" (1654) through "the arithmetization of analysis" (19th century) to "resolution of the four-color conjecture" (1976). But though the mathematical ideas are explained well enough, Eves' book is not based on up-to-date historical scholarship, and errs often in emphasis as well as in detail. For instance, Pascal's religious philosophy, instead of being expounded as related to his mathematics, is seen as "neuroticism"; Newton is supposed to have invented the calculus to solve problems in physics; the history of infinite series jumps from Archimedes to Taylor, with no medieval contributions, much less those of Gregory or the Bernoullis; Cauchy's foundations for the calculus are described as though he didn't know the inequality properties of limits and continuity, while Bolzano's analogous work is not even mentioned; 18th-century formal manipulation of series is termed "a gradual accumulation of absurdities." The book is successful if viewed as a popularization of mathematical ideas, with the historical material used only secondarily, for motivation and for human interest; but it cannot be used as a place to learn how and why the mathematics developed.

To the historian of mathematics, the needs of the mathematical community represent a challenge we should be glad to meet. In fact the past twenty years have seen much good work: important and outstanding research in the history of mathematics, covering all time periods, by people who combine historical and mathematical competence. Let me first indicate my criteria for "good work," and then let me recommend some, beginning with general works and ending with more specialized materials I think are particularly valuable. I hope this essay can serve as a resource for mathematicians, and also galvanize other historians to answer the same questions for the mathematical community from their own points of view.

Works in the history of mathematics should be both mathematically sound and historically correct. Historical correctness does not mean merely having the facts right; it also requires conveying a comprehensive picture of the past and giving evidence that this picture is true. This is easier said than done. The mathematical past is often different from the present: canons of proof are different, attitudes toward what is important are different. Without knowing the past, one cannot—given a modern statement of Newton's approximation or Fermat's theorem—figure out how Newton or Fermat discovered their results, much less how they proved them or what contemporaries thought about the proofs. And the mathematical past often violates our cherished preconceptions. Research in the twentieth century has shown that for instance, the Middle Ages were not mathematically barren; Newton did not invent the calculus solely in order to do his physics; Cayley did not originate the concept of group; Galois did not write all his algebra in one night. Above all, although the mathematical past is causally related to the present, one cannot look at several older papers and infer a causal chain linking them; resemblances do not prove influence. Where there is a causal relationship, historians must demonstrate it, using standards of evidence similar to those which one presents in the natural sciences or in a court of law. Historians, like scientists and lawyers, synthesize their findings from the evidence and conclusions established by their predecessors. It is because of this that recent scholarship in the history of mathematics is worth recommending.

What general histories of mathematics, then, can one recommend as based on up-to-date and reliable historical scholarship? Dirk J. Struik's *Concise History of Mathematics*, 1967, still in print and reasonably priced thanks to Dover, covers mathematics until 1900 in less than two hundred pages, with due attention to social conditions, with a good sense of what is mathematically important, and with chapter bibliographies that include both classic and recent scholarship. For a longer book, Carl Boyer's *History of Mathematics*, 1968 (a reprint at only $12.50 is now available

from Princeton) comprises 700 pages and draws extensively on the scholarly literature up to its publication date. There is a reasonable amount of mathematical content, and each chapter ends with both problems and historically-oriented study questions. Both Struik and Boyer give a good sense of major historical trends in mathematics, and Boyer can serve as a fine course text. Rumor has it that Wiley will be bringing out a new edition of Boyer's book, revised by Uta Merzbach; when this comes out, it will unquestionably be the general history of choice. Morris Kline's 1200-page *Mathematical Thought from Ancient to Modern Times*, 1972, is worth consulting for its detailed accounts of the contents of important mathematical works in the eighteenth and nineteenth centuries.

One will presumably want to read some original mathematical papers, since one's guess at what the mathematicians of the past must have done is invariably more rational and direct than what actually took place. (One's students should read original papers too, since general histories invariably make it all look too easy.) The best collection of sources in English is Dirk Struik's *Source Book in Mathematics, 1200–1800*, 1969; the selections are well chosen and the editing and accompanying explanations are superb. I hope your library has it, because when I wanted it for my course I was told it was available only as a book-on-demand for $130.00. Bring it back, please, Harvard Press. For a course text, then, one must use what is still in print: Ronald Calinger's *Classics of Mathematics*, 1982, the most comprehensive source book yet, which covers work from antiquity to 1932, but whose selections are a bit too short to give an in-depth picture of the nature of the mathematics involved. D. E. Smith's *Source Book in Mathematics*, [1450–1900], 1929, is now out of print. Papers on more modern topics may be found in Garrett Birkhoff's *Source Book in Classical Analysis*, 1973, which focusses on important nineteenth-century work, and in J. van Heijenoort, *From Frege to Goedel: A Source Book in Mathematical Logic, 1879–1931*, 1967. One surely should thank Dover for keeping in print Thomas L. Heath's annotated editions of the Greek mathematicians, especially Euclid and Archimedes; students can sample Babylonian and Greek mathematics in worked-out examples in A. Aaboe's *Episodes from the Early History of Mathematics*, 1964.

The mathematician seeking interesting reading has the most rewarding part of the search yet ahead. Among the excellent recent books in the history of mathematics are Thomas Hawkins's exemplary *Lebesgue's Theory of Integration: Its Origins and Development*, 1970, which conveys much about nineteenth-century mathematics in general; Joseph Dauben's *Georg Cantor*, 1979, which well integrates Cantor's mathematical work with his life and philosophy; Wilbur Knorr's pathbreaking account of Greek mathematicians' ideas of proof and choice of subject-matter, *The Evolution of the Euclidean Elements*, 1975; Michael S. Mahoney's *The Mathematical Career of Pierre de Fermat, 1601–1665*, 1973, explaining Fermat's work and the nature of seventeenth-century mathematics. D. T. Whiteside has done outstanding work on all aspects of Newton's mathematics; his introductions to the eight volumes of the *Mathematical Papers of Isaac Newton* (1967–1981) are highly recommended. Mathematicians will also find interesting A. Rupert Hall, *Philosophers at War: The Quarrel Between Newton and Leibniz*, 1980; Herman Goldstine's *History of Numerical Analysis from the 16th through the 19th Century*, 1977, less for historical narrative than for mathematically sound accounts of the contents of many important individual papers; Jeremy Gray's *Ideas of Space: Euclidean, Non-Euclidean, and Relativistic*, 1979, combining historical and mathematical instruction; Constance Reid's biographies *Hilbert*, 1970; *Courant in Göttingen and New York*, 1976; and *Neyman—From Life*, 1982; each of which conveys much about the mathematical communities in which their subjects lived and worked; Gregory Moore's *Zermelo's Axiom of Choice: Its Origins, Development, and Influence*, 1982; Ian Hacking's *The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction and Statistical Inference*, 1975 ("early" is 1660–1737); and maybe my own *Origins of Cauchy's Rigorous Calculus*, 1981. Carl Boyer's *History of the Calculus*, originally published in 1939 but still available thanks to Dover, his *History of Analytic Geometry*, 1956, and Otto Neugebauer's *Exact Sciences in Antiquity*, 1957, are, though older, still worth reading.

It is well worth sampling the periodical literature, also, especially *Historia Mathematica* (which

reviews books in the history of mathematics, so that the reader can continue to update the list in this essay), and the *Archive for History of Exact Sciences*. The equivalent of Bartlett's *Quotations* in the history of mathematics is R. E. Moritz, *On Mathematics and Mathematicians* (originally published under the title *Memorabilia Mathematica*, in 1914; the Dover Edition is now out of print); the equivalent of *Mathematical Reviews* is some combination of the annual Critical Bibliography issue of *Isis* (the journal of the History of Science Society), the appropriate sections of *Math. Reviews*, and the Abstracts section of *Historia Mathematica*. For an excellent, critically-annotated bibliography of 2,384 items, there is the just-published *History of Mathematics from Antiquity to the Present: A Selective Bibliography*, ed. J. W. Dauben, Garland Press, 1985 (every library should have this!); and the equivalent of *Who Was Who* (i.e., biographies of mathematicians of the past who are deceased) is the *Dictionary of Scientific Biography*, 1970– , (especially recommended: H. Freudenthal on Cauchy, A. P. Yushkevich on Euler). Last but not least, there are classics in the popularization of mathematical ideas and the nature of the mathematician's profession which, by virtue of the outstanding mathematicians who have written them, are worth attention, especially for students who want to figure out what "mathematics" and "doing mathematics" mean. Jacques Hadamard's *Psychology of Invention in the Mathematical Field*, 1945, G. H. Hardy's *A Mathematician's Apology*, 1940, and Reuben Hersh and Philip Davis, *The Mathematical Experience*, 1980—all in paperback—greatly enrich a course's reading list.

In sampling the historical literature of the past twenty years of interest to mathematicians, I have of course omitted some interesting items, and have also omitted references to the periodical literature, even though the articles are quite important. How would the reader of this essay ever find out about Thomas Hawkins' papers in the *Archive* on nineteenth-century mathematics, or the writings of Thomas Hankins and Helena Pycior on Hamilton, or Lorraine Daston's work on probability, or Joan L. Richards' work on non-Euclidean geometry? And where should one include such interesting but hard-to-classify works as the English version of Sofya Kovalevskaya's *Russian Childhood*, 1978; Claudia Zaslavsky's *Africa Counts*, 1973, or Joseph Weizenbaum's superb *Computer Power and Human Reason*, 1976? Still, there is a great deal here. When I began my graduate work over twenty years ago, only Struik's *Concise History*, the Smith *Source Book*, Boyer's *Calculus* and *Analytic Geometry*, Heath's editions of the Greeks, and Neugebauer's *Exact Sciences in Antiquity* were easily available. Neither the *Archive for History of Exact Sciences* nor *Historia Mathematica* existed. Since then, much new work has appeared; some of it is truly outstanding. We live in an age of riches; we should take advantage of it.

Note added in proof. I have just learned that the Princeton University Press will reprint Struik's *Source Book in Mathematics, 1200–1800*, at about $12.50 in paper. Princeton and Dover deserve the thanks of the mathematical community for reprinting Boyer, Struik, Euclid, and other classics at prices students can afford.

Notes on Introductory Combinatorics. By George Pólya, Robert E. Tarjan, and Donald R. Woods. Birkhauser, Boston, 1983. 190 pp.

A. Wayne Roberts
Mathematics and Computer Science Department, Macalester College, St. Paul, MN 55105

Popular characterizations notwithstanding, most of us have known good research mathematicians who have also been interested and effective in teaching. Among first rank research mathematicians, however, very few manage to establish beyond the borders of their own schools a reputation for teaching. George Pólya did. Most of us who ponder how best to teach mathematics have at one time or another tried *Let's Teach Guessing* as one approach to *How to Solve It*.

For these reasons, this set of Notes will be of particular interest. Compiled by teaching

assistant Donald Woods in a course taught by Pólya and Robert Tarjan, these are different (and better) than the usual notes in which a well-intentioned scribe of the Gaussian school has carefully removed the scaffolding before exposing to us an austere collection of definitions, theorems and proofs. These notes preserve the dynamics of the classroom, the motivational techniques, even the sense of timing that lets us in on what questions were posed just before a class ended. It is true, of course, that little jests offered by the then 90-year old Pólya probably came across better in the classroom than they do in writing, but even to this level of detail, Woods has let us all see just how Pólya conducted a class.

So when we get a glimpse inside his classroom, what do we see? ABRACADABRA. We see an exploration of how many ways there are to spell it out, always going from one letter to an adjacent letter in the arrangement

```
          A
         B B
        R R R
       A A A A
      C C C C C
     A A A A A A
      D D D D D
       A A A A
        B B B
         R R
          A
```

We are told that the class decided by vote that the number exceeded 100, and are led to believe that the vote for more than 1000 was too close to call. And we see that Pólya seized classroom opportunities to enunciate such problem solving maxims as "If you cannot solve the proposed problem, solve first a suitable related problem." Acting on this principle, Pólya soon had Pascal's triangle worked out.

One is struck throughout the notes with Pólya's willingness to risk being didactic, to put into words the general principles that we perhaps too frequently suppose students will notice without our stating them in so many words. Time and again Pólya pauses to tell students such things as,

> "The beginning of most discoveries is to recognize a pattern." (page 6)
> "Start by working out the first few 'small' cases, and look for a pattern." (page 30)
> "If you see a fact, try to see it as intuitively as possible." (page 70)

How would our teaching be affected if, as class proceeded, we kept a running list on the side board of each principle illustrated by our activities in front of the class that day?

One is also struck with Pólya's willingness to take a good deal of time on details. There are several accounts in which he used precious class time for calculations needed to fill in a couple of rows of a table. I counted four instances where, having obtained a general formula, he took class time to go back and confirm that it worked for the first few cases that had been used to introduce the topic (thus carrying out another of his maxims). He was fond of returning to a solved problem, looking at it again to draw out another lesson, or raising an apparently idle question. Having obtained Pascal's triangle, he casually noted that for some values of n, all but the first and last entries of row n are divisible by n, and from this observation came Fermat's Little Theorem. He next got to wondering how many numbers in row n are odd. Before he was through with that, students had been introduced to binary representation of numbers and had been given another lesson on the value of noticing patterns.

Attention to details need not obscure the larger picture. Pólya's genius for giving the large picture through the use of specific examples is beautifully illustrated in his discussion of *counting* combinatorics, *existential* combinatorics, and *constructive* combinatorics (pp. 86–87).

Rigor was often sacrificed to insight, to getting a feel for a result. This is seen everywhere but is striking in his handling of what is commonly called Pólya's Theory of Counting. Even when the work was his own, he was able to resist giving proofs. Wanting to use a theorem, the proof of which was considered to be beyond the scope of the course he remarked, "The proof of the pudding is in the eating. You can't eat mathematics, but you can digest it. So let's digest this theorem by chewing on a few examples."

Our focus on Pólya is not a negative reflection on Tarjan who taught the second half of the course. Indeed, the notes suggest that Tarjan was a worthy co-teacher who, while moving the course along to standard results in algorithmic graph theory (Ramsey Theory, Matchings, Network Flow), also used the technique of motivating with specific problems; and he stated a few maxims of his own.

Nevertheless, confined as it is to exactly those topics chosen in a course of standard content in combinatorics, it is not likely that other teachers will find here all the topics they would want to include in their course. This book deserves to be known because of the snapshot it gives us of Pólya in the classroom.

LETTERS TO THE EDITOR

For instructions about submitting letters for publication in this department see the inside front cover.

Editor:

In a recent article in this MONTHLY [**1**], A. Abian studied the solvability of infinite systems of polynomial equations over the field $\mathbb{C}$ of complex numbers. For ease of reference, we restate Abian's main result [**1**, Theorem and Remark 3]:

Let K be an algebraically closed field and D a system of polynomial equations over K with $|D| < |K|$. Then D has a solution in K if and only if every finite subsystem of D has a solution in K.

Abian's result may be used to give a straightforward proof of the following theorem, which generalizes to infinitely many variables the following weak form of the celebrated Hilbert Nullstellensatz [**2**, Theorem 32]:

If K is an algebraically closed field, then any maximal ideal M in the polynomial ring $R = K[X_1,\dots,X_n]$ is of the form $(X_1 - a_1,\dots,X_n - a_n)$ for some $a_i \in K$, i.e., it consists of all polynomials vanishing at a point.

THEOREM. *Let K be a field and $\{X_\alpha | \alpha \in \Lambda\}$ a family of indeterminates over K. Then every maximal ideal of the polynomial ring $R = K[\{X_\alpha|\alpha \in \Lambda\}]$ has the form $(\{X_\alpha - a_\alpha | \alpha \in \Lambda\})$ for some $a_\alpha \in K$ if and only if K is algebraically closed and $|\Lambda| < |K|$.*

Proof. First, assume that K is algebraically closed and $|\Lambda| < |K|$. Let M be a maximal ideal of R. Then $M = (D)$, where $|D| = |\Lambda| < |K|$ (for Λ finite, c.f. [**2**, Exercise 6, page 20], while for Λ infinite, this follows because $M = \cup(M \cap K[X_{\alpha_1},\dots,X_{\alpha_n}])$, where this union is over all finite subsets of Λ, and each $M \cap K[X_{\alpha_1},\dots,X_{\alpha_n}]$ is finitely generated by the Hilbert Basis Theorem [**2**, Theorem 69]). Let D' be a finite subset of D and let $\{X_{\alpha_1},\dots,X_{\alpha_n}\}$ be the unknowns that appear in the polynomials in D'. Then $D' \subset M \cap K[X_{\alpha_1},\dots,X_{\alpha_n}]$, which is a proper prime ideal of $R' = K[X_{\alpha_1},\dots,X_{\alpha_n}]$. By the weak form of the Nullstellensatz, the polynomials in D' have a common zero in K. By Abian's result, the polynomials in D also have a common zero in K. Thus $M \subset (\{X_\alpha - a_\alpha | \alpha \in \Lambda\})$ for some $a_\alpha \in K$, and hence $M = (\{X_\alpha - a_\alpha|\alpha \in \Lambda\})$.

Conversely, assume that either K is not algebraically closed or $|\Lambda| \geqslant |K|$. If K is not

algebraically closed, then one may easily construct a maximal ideal not of the desired form. Hence we assume that K is algebraically closed and $|\Lambda| \geqslant |K|$, and thus Λ is infinite. Choose a fixed $X_0 \in \{X_\alpha | \alpha \in \Lambda\}$, and for each $a_\beta \in K$ choose distinct $X_\beta \in \{X_\alpha | \alpha \in \Lambda\} - \{X_0\}$. Then let

$$D = \{(X_0 - a_\beta)X_\beta - 1 | a_\beta \in K\}.$$

Clearly each finite subset of D has a common zero in K, but D itself has no common zero in K. Thus (D) is contained in some maximal ideal M of R (for if not, then $g_1 f_1 + \cdots + g_n f_n = 1$ for some $g_i \in R$ and $f_i \in D$, and upon substituting the guaranteed common zero for the finite subset $\{f_1, \ldots, f_n\}$ of D, we obtain the contradiction $0 = 1$), and M is *not* of the form $(\{X_\alpha - a_\alpha | \alpha \in \Lambda\})$ for any $a_\alpha \in K$. ■

Note that our theorem includes the above-mentioned weak form of the Nullstellensatz since any algebraically closed field is infinite. Although we will not prove this, it may be easily seen that our theorem is actually equivalent to Abian's result. Finally, note that our theorem shows that when considering the solvability of a system D of polynomial equations, the important thing is the number of unknowns that appears in D, and not $|D|$ (note, however, that $|\text{unknowns}| \leqslant |D|$, for D infinite).

References

1. A. Abian, Solvability of infinite systems of polynomial equations over the field of complex numbers, this MONTHLY, 92 (1985) 94–98.

2. I. Kaplansky, Commutative Rings, revised edition, The University of Chicago Press, Chicago, 1974.

David F. Anderson
Department of Mathematics
The University of Tennessee
Knoxville, TN 37996

Editor:

In his note "Dini type theorems on positive series" (this MONTHLY, 92(1985), 347–349) Piotr Biler proves (Theorem 2) that each of two conditions on a positive function F on R^+ is equivalent to the convergence of $\Sigma a_n F(s_n)$ for all positive divergent series Σa_n, where $s_n = \sum_1^n a_m$. One of these equivalences, that there exists a decreasing function h, integrable on $(\delta, \infty)_1$, $\delta > 0$, that majorizes F ($F \leqslant h$), is essentially the same as Theorem 2 of my generalized solution to problem E 2558 (1975, 936) "A theorem of Dini" (this MONTHLY, 84(1977), 138–139) where it was shown that for $g(x) \geqslant 0$, $x \geqslant 0$, and $h(x) = \sup_{y \geqslant x} g(y) \leqslant \infty$, the series $\Sigma a_n g(s_n)$ converges for all positive divergent series Σa_n if and only if there exists $\alpha \geqslant 0$ such that $\int_\alpha^\infty h(x)\,dx < \infty$.

Biler's Theorem 1 considers the analogous problem for convergent positive series Σa_n and he proves that, for positive F, the series $\Sigma a_n F(r_n)$ will always converge (here r_n is the tail $\sum_n^\infty a_m$) if and only if there exist a $\delta > 0$ and a decreasing function h on $(0, \delta)$ such that h is integrable on $(0, \delta)$ and h majorizes F ($F \leqslant h$).

Biler fails, however, to consider the naturally suggested problem, in the case of positive divergent Σa_n, of what conditions on $F > 0$ are equivalent to the *divergence* of $\Sigma a_n F(s_n)$ for all such series: this was answered by Theorem 1 of "A theorem of Dini" where it was proved that for

$f(x) > 0$ for $x > 0$ the series $\Sigma a_n/f(s_n)$ diverges for all positive divergent series Σa_n if and only if $f(x) = O(x)$ as $x \to \infty$.

Neither paper considers the remaining case of convergent Σa_n and divergent $\Sigma a_n F(r_n)$, so I would like to complete the whole discussion by the following result.

THEOREM. *Let $f(x) > 0$ for $x > 0$. Then $\Sigma a_n/f(r_n) = \infty$ for every positive convergent series Σa_n if and only if $f(x) = O(x)$ as $x \to 0+$.*

Proof. If $f(x) < Ax$ for $0 < x < x_0$, and $r_k < x_0$, then

$$\Sigma a_n/f(r_n) \geqslant A^{-1} \sum_{n \geqslant k} a_n/r_n = \infty,$$

by the classical Dini theorem that $\Sigma a_n/r_n = \infty$.

If $f(x) \neq O(x)$ as $x \to 0+$, then $\liminf x/f(x) = 0$ as $x \to 0+$ and hence one may define $r_1 > r_2 > \cdots (r_n \downarrow 0)$, so that, with $a_n = r_n - r_{n+1}$,

$$a_n/f(r_n) = \frac{(r_n - r_{n+1})}{r_n} \frac{r_n}{f(r_n)} < 2^{-n}.$$

Then $\Sigma a_n = r_1$ converges but $\Sigma a_n/f(r_n) < \infty$.

M. J. Pelling
Department of Mathematics
University College
London WClE 6BT, England

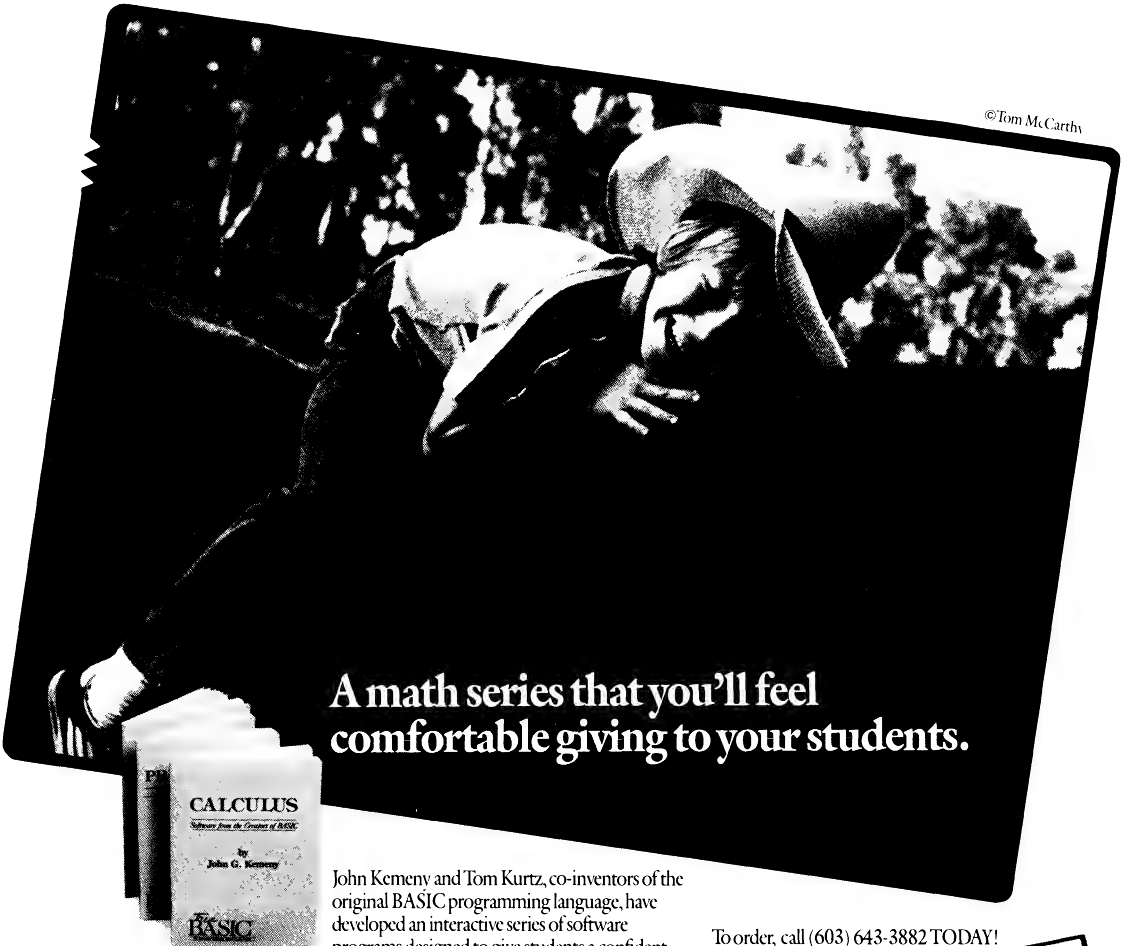
---

171. MISCELLANEA

Excelsior!

Peter Popponesset was as brilliant as could be,
He wouldn't work on any petty problems, no, not he!
His prof suggested topics; Peter said they were too soft,
For he would hold the flag of great significance aloft.
He scorned Riemann's conjecture—he claimed it was passé:
He wouldn't waste his time on classic stumbling blocks today.
He's watching and he's waiting for a challenge to arise
Whose answer will provide sufficient glory in his eyes.
His classmates were pedestrian; they got their Ph.D.'s;
They solved some open problems, and promotions came with ease.
They've even won some prizes given by the A.M.S.,
But Peter sneers at all of them and mocks at their "success."
He knows that he'll laugh last, not now, but in the end,
Because he's still expecting inspiration to descend.

Ralph P. Boas
Department of Mathematics
Northwestern University
Evanston, IL 60201



John Kemeny and Tom Kurtz, co-inventors of the original BASIC programming language, have developed an interactive series of software programs designed to give students a confident understanding of the principles of mathematics.

These programs include Pre-Calculus, Calculus, Algebra, Trigonometry, Probability Theory, Statistics, and Discrete Mathematics.

The True BASIC™ Mathematics Series will fit into any teacher's individual classroom program. They're not tied to any particular text or course. They supplement your own curriculum at your own speed.

They're reasonably priced at $49.95. Generous quantity discounts are available.

The True BASIC Mathematics Series. You can be confident using it as the workhorse in your classroom. It works comfortably with the IBM-PC™ and compatibles, Apple Macintosh,™ and Commodore Amiga.™

To order, call (603) 643-3882 TODAY!

Or send the coupon below to receive a FREE demo disk.

FREE DEMO DISK OFFER.

True BASIC™ inc.
39 South Main Street,
Hanover, NH 03755 (603) 643-3882

My computer is: ☐ IBM-PC/compatible
☐ Apple Macintosh
☐ Commodore Amiga

☐ I'D LIKE TO GET MY FREE DEMO DISK.
☐ I'd like more information on all of the mathematics products.
☐ I'm ready to buy. Call me and tell me how.
☐ I'd like information on True BASIC Language products.

Name ______
Title ______
Company/University ______
Address ______
City, State, ZIP ______
Telephone ______

AMM-7

DIRECTORY OF WOMEN IN THE MATHEMATICAL SCIENCES

The Association for Women in Mathematics plans to publish a new edition of the *Directory of Women in the Mathematical Sciences* to bring up to date the earlier directories published by AMS and by the AMS-MAA-NCTM-SIAM Committee on Women in Mathematics.

The purpose of the *Directory* is to make available, in summary form, curricula vitae of women in the mathematical sciences for employers, organizers of colloquia and conferences, and officers of national organizations.

Women who hold the Ph.D., have the equivalent training, or who are Ph.D. candidates, and who wish to be listed in the new *Directory* should complete the questionaire below and mail it to the address at the bottom of this page.

(PLEASE TYPE; handwritten forms may not be used.)

1. A. Name (last name first) ____________________

   B. Cross listing of name (if desired) ____________________

2. Mailing Address (Preferred) ____________________

____

____

*Current Status ____ (Employed) ____ (Retired) ____ (Graduate Student) ____ (Unemployed)
(*This information is for record keeping and will not be printed in the Directory.)

3. Name of employer or institution ____________________
   (If retired, give last institution where employed.)

   Address of employer or institution ____________________

   ____

   Title of position ____________________

4. Highest academic degree received ____________________

   Year degree conferred ________ Institution ____________________

5. Fields of mathematical interest: (Write 1 for primary field; 2 for other. Only two will be printed.)

A. ____ Algebra
B. ____ Analysis
C. ____ Applied mathematics
D. ____ Combinatorics
E. ____ Computer science
F. ____ Functional analysis
G. ____ Geometry
H. ____ Logic and Foundations
I. ____ Number theory
J. ____ Operations research
K. ____ Statistics and probability
L. ____ Topology
M. ____ Other

6. Publications: Bibliographic information for two most important publications. In addition to the title, give MR number and date only, if available; otherwise for an article, use MR abbreviation for journal, give volume number and date.

   i. ____________________

   ____

   ____

   ____

   ii. ____________________

   ____

   ____

   ____

Return this form as soon as possible to Directory of Women in the Mathematical Sciences, c/o R. Struik, Math Dept., Campus Box 426, University of Colorado, Boulder, CO 80309.

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 93, Number 7 August-September 1986

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. (This is necessary for indexing purposes.) Send all main articles, letters to the editor, and editorial correspondence to the editor-elect: HERBERT S. WILF, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Contributions to the other four special departments (Notes, Teaching of Mathematics, Unsolved Problems, and Problems and Solutions) should be sent to the appropriate one of the associate editors, as follows:

Notes: CAROL G. CRAWFORD, Department of Mathematics, U.S. Naval Academy, Annapolis, MD 21402, *or* RICHARD LIBERA, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716, *or* ANITA E. SOLOW, Department of Mathematics, Grinnell College, Grinnell, IA 50112.

Teaching of Mathematics: JOAN P. HUTCHINSON *or* STAN WAGON, Department of Mathematics, Smith College, Northampton, MA 01063.

Unsolved Problems: RICHARD GUY, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.

Proposed problems and solutions, both elementary and advanced: PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801.

Reprint Permission: A. B. WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Back Issues: P. and H. BLISS Co., Middletown, CT 06457.

Microfilm Editions: University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

Editorial Assistants: Fanny Hailpern, Mary P. Jackson.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT, ten times a year: January, February, March, April, May, June-July, August-September, October, November and December.

The annual subscription price for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is $24 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $24. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $70 per year.

Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036

PRINTED IN THE UNITED STATES OF AMERICA

LUDWIG BIEBERBACH'S CONJECTURE AND ITS PROOF BY LOUIS DE BRANGES

J. KOREVAAR

Mathematics Institute, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands

Summary. 1984 has been an exciting year for complex analysis. It even brought strong rumors that the Riemann hypothesis had been proved, but so far, the rumor has not been confirmed. However, we know for sure that the difficult Bieberbach conjecture has been settled this year. As many of you know, this famous conjecture of 1916 concerns the class S of normalized injective holomorphic functions. That class consists of the 1-1 holomorphic functions from the unit disc U into the complex plane $\mathbb{C}$ with a power series of the form

$$f(z) = z + a_2 z^2 + \cdots + a_n z^n + \cdots, \quad |z| < 1.$$

The conjecture asserts that $|a_n| \leqslant n$ for every f in S and every n. Louis de Branges has proved this conjecture as well as some stronger conjectures for the class S.

Each of the following items has played an essential role in the proof:

(i) Löwner's partial differential equation for so-called Löwner chains $\{f_t(z)\}$ of injective holomorphic functions from U to $\mathbb{C}$.

(ii) The observations of Lebedev and I. M. Milin, especially their inspired conjecture for the so-called logarithmic coefficients of f in S, that is, the coefficients in the expansion $\sum_1^\infty c_k z^k$ for a branch of $\log\{f(z)/z\}$.

(iii) De Branges' striking breakthrough, namely, the creation of a functional associated with the Lebedev-Milin conjecture which varies monotonically along Löwner chains.

(iv) De Branges' introduction and solution of a system of differential equations which he devised to make the functional manageable.

(v) A positivity result for hypergeometric functions which is a tool in establishing the monotonicity of the functional.

Of the above, (i) dates back to 1923, while (ii) and (v) are relatively recent. The Lebedev-Milin observations date from the years 1965–1970 and became well known in the West only around 1977. The hypergeometric functions result occurs in work of Askey and Gasper of 1976.

1. Historical introduction. Our starting point is the well-known *conformal mapping theorem* formulated by Riemann. Let D be an arbitrary simply connected domain in the complex plane $\mathbb{C}$ which is not the whole plane. Then there exists a conformal (or 1-1 holomorphic) map $w = f(z)$ from the unit disc U: $\{|z| < 1\}$ onto D. One may arbitrarily prescribe the image $f(0)$ (in D) of the origin, as well as the angle $\arg f'(0)$ through which directions at the origin are rotated. However, such data determine the map uniquely. The first complete proofs of the theorem were given around 1900: by Hilbert, who put the Dirichlet principle on a rigorous basis, and by Osgood, who constructed and used Green's functions for D.

Questions on the fine structure of conformal maps became a popular topic in German mathematics around 1910 (Koebe, Carathéodory and others). Let us normalize our injective holomorphic maps f from U to $\mathbb{C}$ by requiring $f(0) = 0$, $f'(0) = 1$. Then we obtain the class S of normalized 1-1 holomorphic functions ("schlicht" or univalent functions)

$$f(z) = z + a_2 z^2 + \cdots + a_n z^n + \cdots \text{ on } U. \tag{1}$$

The author has worked in complex analysis and approximation theory, including Tauberian theorems and Müntz-type approximation. Born and educated in the Netherlands, he became an admirer of Hardy-Littlewood-Pólya-Szegö-Wiener. He has spent 25 years in the U.S., mostly at the University of Wisconsin (Madison) and the University of California (San Diego). Since 1974 he has been a professor at the University of Amsterdam and a member of the Netherlands Academy of Sciences; he holds an honorary degree from the University of Gothenburg. Mathematics is a family interest (the father-in-law story applies), as are music, languages and mountain hiking.

EXAMPLES. The formula

$$w = \frac{1+z}{1-z} = 1 + 2z + \cdots + 2z^n + \cdots$$

defines a conformal map of U onto the right half-plane H: $\{\operatorname{Re} w > 0\}$. Squaring, we obtain a conformal map

$$w = \left(\frac{1+z}{1-z}\right)^2$$

of U onto the slit plane $\mathbb{C} \setminus (-\infty, 0]$. Normalization gives the important *Koebe function*

$$K_0(z) = \frac{1}{4}\left\{\left(\frac{1+z}{1-z}\right)^2 - 1\right\} = \frac{z}{(1-z)^2} = z + 2z^2 + \cdots + nz^n + \cdots \tag{2a}$$

which maps U onto $\mathbb{C} \setminus (-\infty, -\frac{1}{4}]$. This function and its rotations

$$K_\theta(z) = e^{-i\theta} K_0(e^{i\theta} z) = \frac{z}{(1 - e^{i\theta} z)^2} \tag{2b}$$

(which we will also call Koebe functions) provide the solution to many extremal problems for the class S.

Some extremal problems. In 1916 Bieberbach **[2]** proved that $|a_2| \leqslant 2$ for every f in S, with equality only for the Koebe functions (2). In a footnote he remarked that perhaps quite generally

$$|a_n| \leqslant n \quad \text{for } f \in S. \tag{3}$$

This footnote became the famous Bieberbach conjecture which remained unproven until 1985, although a great deal of work was expended on it. The a_2-result can be used to show that the image $f(U)$ contains the disc $|w| < 1/4$ in the w-plane for every f in S. Moreover, if $f(U)$ contains no larger disc about 0, then f is a Koebe function. Related results are Koebe's distortion theorems, of which we mention

$$|f'(z)| \leqslant \frac{1+|z|}{(1-|z|)^3}, \quad \frac{|z|}{(1+|z|)^2} \leqslant |f(z)| \leqslant \frac{|z|}{(1-|z|)^2}. \tag{4}$$

These inequalities hold for all f in S, with strict inequality for all $z \neq 0$ unless f is a Koebe function.

General references for this section and the next are the books by Goluzin **[10]**, Pommerenke **[14]**, Duren **[6]** and Goodman **[11]**.

2. More on the Bieberbach conjecture. Using the partial differential equation named after him, see Section 4(iii), Löwner proved in 1923 that $|a_3| \leqslant 3$ for every f in S **[12]**. Later, Schiffer and others developed a number of variational methods for injective holomorphic functions. In the years 1955–1972 those techniques yielded rather laborious proofs for the special cases $n = 4$, 6 and 5 of the Bieberbach conjecture. From time to time, proofs for other special cases have been announced, but they have not been substantiated.

Turning to general n, the upper bound for $|f(z)|$ in the distortion relations (4) and Cauchy's inequality for the coefficients of a power series readily show that $|a_n| < en^2$. In 1925 Littlewood found the correct order of the upper bound for $|a_n|$ as $n \to \infty$:

$$|a_n| < en \quad \text{for all } f \in S.$$

The best result of this kind until this year was that of FitzGerald (1972), including a slight improvement by his student Horowitz (1978):

$$|a_n| < 1.07n.$$

There is also a beautiful regularity theorem of Hayman (1953): the limit

$$\lim_{n\to\infty} \frac{|a_n|}{n}$$

exists for every f in S, and is smaller than 1 unless f is a Koebe function. This result remains of interest even now that the Bieberbach conjecture has been proved.

Odd functions. For f in S, it is often useful to look at the related function

$$f_1(z) = \sqrt{f(z^2)} = b_1 z + b_3 z^3 + \cdots + b_{2n-1} z^{2n-1} + \cdots, \quad b_1 = 1. \tag{5}$$

This is an odd function in S, and every odd function f_1 in S can be represented as such a square root transform. By (1) and (5)

$$a_n = b_1 b_{2n-1} + \cdots + b_{2n-1} b_1. \tag{6}$$

If f is the Koebe function K_0 then

$$f_1(z) = \frac{z}{1-z^2} = z + z^3 + \cdots + z^{2n-1} + \cdots.$$

In 1932 Littlewood and Paley proved that there is a constant $C \leqslant 14$ such that for all odd functions in S,

$$|b_{2n-1}| \leqslant C, \quad n = 1, 2, \ldots.$$

In a footnote they remarked: "No doubt the true bound is given by $C = 1$." Observe that by (6), the truth of this conjecture would imply the Bieberbach conjecture!

Recently V. I. Milin proved that one may take $C = 1.14$. Before that, Hayman had shown that

$$\lim_{n\to\infty} |b_{2n-1}|$$

exists for every odd function f_1 in S, and is smaller than 1 unless f_1 is the root transform of a Koebe function. These results would seem to support the Littlewood-Paley conjecture. However, the latter had been disproved already in 1933 by Fekete and Szegö: there exist odd functions in S for which $|b_5| > 1$. No wonder that some experts doubted the Bieberbach conjecture as well!

However, there are always people around with the intuition to come up with a good conjecture. It was observed by Robertson that by (6) and the Cauchy-Schwarz inequality, the Bieberbach conjecture would already follow from the inequality

$$\sum_1^n |b_{2k-1}|^2 \leqslant n. \tag{7}$$

(7) became known as the *Robertson conjecture* for odd functions in S (1936).

Logarithmic coefficients. Since 1940, one has increasingly used certain logarithmic transforms of injective holomorphic functions. The associated Grunsky and Goluzin inequalities have been successfully applied to various extremal problems. More recently, Lebedev and I. M. Milin have focused on the expansion

$$\log \frac{f(z)}{z} = \sum_1^\infty c_k z^k, \quad |z| < 1 \tag{8}$$

for f in S. (Note that $f(z)/z$ is holomorphic and zero-free on U; one takes the branch of the logarithm which vanishes at the origin.) For the Koebe function K_0 one has $c_k = 2/k$. For the case of image domains $f(U)$ that are star-shaped relative to the origin, one has $|c_k| \leqslant 2/k$ and this inequality readily implies the Bieberbach conjecture (3) for such "starlike" functions f.

The latter result goes back to Nevanlinna (1920). For starlike f, a geometric argument at the boundary shows that

$$\operatorname{Re} z\frac{f'(z)}{f(z)} = \operatorname{Re}\left(1 + \sum_1^\infty kc_k z^k\right) > 0,$$

so that by a well-known inequality of Carathéodory for functions with positive real part, $k|c_k| \leqslant 2$, cf. Section 4(vi).

The inequality $|c_k| \leqslant 2/k$ or

$$k|c_k|^2 - \frac{4}{k} \leqslant 0$$

is not true for every f in S, but Lebedev and Milin conjectured that the latter inequality is true in the following average sense:

$$\Omega_n \overset{\text{def}}{=} \sum_{p=1}^{n-1} \sum_{k=1}^{p} \left(k|c_k|^2 - \frac{4}{k}\right) = \sum_{k=1}^{n-1}\left(k|c_k|^2 - \frac{4}{k}\right)(n-k) \leqslant 0 \tag{9}$$

for $n = 2, 3, \ldots$ and all $f \in S$. This amazing conjecture occurs in Milin's book of 1971; the book became available in English only in 1977 [**13**]. The L-M conjecture implies the Robertson conjecture and hence also the Bieberbach conjecture.

Examples. The L-M conjecture is easy to prove for $n = 2$ and 3, cf. Section 4(vii). For $n = 2$ it asserts that $|c_1|^2 \leqslant 4$. Since

$$\begin{aligned} 1 + a_2 z + a_3 z^2 + \cdots = \frac{f(z)}{z} &= \exp(c_1 z + c_2 z^2 + \cdots) \\ &= 1 + c_1 z + \left(\frac{1}{2}c_1^2 + c_2\right) z^2 + \cdots, \end{aligned} \tag{10}$$

Bieberbach's inequality is an immediate consequence:

$$|a_2| = |c_1| \leqslant 2.$$

For $n = 3$ the L-M conjecture is equivalent to $|c_1|^2 + |c_2|^2 \leqslant 5$. Löwner's inequality is an easy corollary:

$$|a_3| = \left|\frac{1}{2}c_1^2 + c_2\right| \leqslant \frac{1}{2}|c_1|^2 + |c_2| \leqslant \frac{5}{2} - \frac{1}{2}|c_2|^2 + |c_2| \leqslant 3$$

(by calculus!).

Turning to the general case, there is a useful inequality of Lebedev and Milin for the coefficients of

$$\sum_0^\infty \beta_k z^k = \exp\left(\sum_1^\infty \gamma_k z^k\right).$$

It asserts that

$$\sum_0^{n-1} |\beta_k|^2 \leqslant n \exp\left\{\frac{1}{n}\sum_{p=1}^{n-1}\sum_{k=1}^{p}\left(k|\gamma_k|^2 - \frac{1}{k}\right)\right\}, \quad n = 1, 2, \ldots,$$

see [**6**]. Applying this inequality to the identity

$$\begin{aligned} b_1 + b_3 z + \cdots + b_{2n-1} z^{n-1} + \cdots &= \frac{f_1(z^{\frac{1}{2}})}{z^{\frac{1}{2}}} \\ &= \left\{\frac{f(z)}{z}\right\}^{\frac{1}{2}} = \exp\left\{\frac{1}{2}\log\frac{f(z)}{z}\right\} = \exp\left(\frac{1}{2}\sum_1^\infty c_k z^k\right), \end{aligned}$$

cf. (5) and (8), one obtains

$$|b_1|^2 + \cdots + |b_{2n-1}|^2 \leqslant n \exp(\Omega_n/4n). \tag{11}$$

Thus if the L-M conjecture (9) holds for $f \in S$ and a certain n, then the Robertson conjecture (7) holds for the corresponding f_1 and the same n, so that also the Bieberbach conjecture (3) must be true for the same n, cf. (6). Moreover, if $\Omega_n < 0$ for some n, then one has strict inequality in (7) and hence also in (3).

De Branges has proved the L-M conjecture and thereby also the Robertson and Bieberbach conjectures [3], [4], [5].

3. De Branges' theorem [4]. *Let f be an arbitrary function in S, let the power-series coefficients a_n be defined by* (1) *and the logarithmic coefficients c_k by* (8). *Then the conjectured* L-M *inequality* (9) *and hence the conjectured Bieberbach inequality* (3) *are true for every $n \geqslant 1$. Equality holds in* (3) *and hence in* (9) *for a certain $n \geqslant 2$ if and only if f is a Koebe function* (2).

4. The proof. I will present de Branges' ideas in as simple a way as I can. The following arrangement is based on de Branges' lecture in Amsterdam (July 10) and an early write-up in Russian which he had at that time. It also shows the influence of his later manuscript [4] and of a manuscript by FitzGerald and Pommerenke [8] based on de Branges' work. The proof will be spread over a number of steps.

(i) *We may take D nice.* For the proof of the L-M conjecture (9), it may be assumed that f maps U onto a domain D bounded by an analytic Jordan curve. Indeed, for any given f in S and $0 < \rho < 1$ we may define

$$f^*(z) = \frac{1}{\rho} f(\rho z) = z + a_2 \rho z^2 + \cdots + a_n \rho^{n-1} z^n + \cdots.$$

The function f^* maps U onto the set $(1/\rho)f(\rho U)$. The latter domain is bounded by the analytic Jordan curve given by $(1/\rho)$ times the image of the circle $|z| = \rho$ under f.

Since

$$\log \frac{f^*(z)}{z} = \log \frac{f(\rho z)}{\rho z} = \sum_1^\infty c_k \rho^k z^k,$$

the logarithmic coefficients c_k^* for f^* are equal to $c_k \rho^k$. Hence if (9) has been proved for the coefficients c_k^*, it follows for the coefficients c_k by letting ρ tend to 1.

(ii) *Löwner chains.* Given $D = f(U)$ as in (i), it is easy to construct a nice continuously increasing family of simply connected domains D_t, $0 \leqslant t < \infty$, such that

$$D_0 = D, \quad D_s \subsetneqq D_t \text{ if } s < t \quad \text{and} \quad D_t \to \mathbb{C} \text{ as } t \to \infty. \tag{12a}$$

One can actually do this for *every* simply connected domain D, cf. [14].

We define

$$f_t(z) = f(z, t), \quad 0 \leqslant t < \infty,$$

as the 1-1 conformal map of U onto D_t such that

$$f_t(0) = 0, \quad f_t'(0) > 0.$$

Then $\omega(t) = f_t'(0)$ will be a strictly increasing continuous function such that $\omega(0) = 1$ and $\omega(t) \to \infty$ as $t \to \infty$. Introducing a new parameter u by setting $\omega(t) = e^u$, if necessary, one may assume from the beginning that $\omega(t) = e^t$. The corresponding family of injective holomorphic functions

$$f_t(z) = f(z, t) = e^t(z + a_2(t) z^2 + \cdots), \quad 0 \leqslant t < \infty; \quad f_0(z) = f(z) \tag{12b}$$

(which depend continuously on t) is called a Löwner chain starting at $f(z)$.

A little more effort shows that *every* $f \in S$ is the starting point of a Löwner chain, cf. [**14**] p. 159.

(iii) *Heuristic derivation of the Löwner differential equation*. The functions $f(z, t)$ of a Löwner chain satisfy the partial differential equation of Löwner [**12**]:

$$\frac{\partial f}{\partial t} = z\frac{\partial f}{\partial z}p(z, t), \tag{13a}$$

where

$$p(z, t) \text{ is analytic in } z, \quad \operatorname{Re} p(z, t) > 0, \quad p(0, t) = 1. \tag{13b}$$

Geometrically, equation (13) represents an outward flow in the plane. Indeed, the vector z gives the direction of the outward normal to the circle C_r: $|z| = r$. Thus $z(\partial f/\partial z)$ gives the direction of the outward normal to the curve $f(C_r)$ at the point $f(z, t)$. By (13), the velocity vector $\partial f/\partial t$ should make an angle with the normal there less than $\frac{1}{2}\pi$.

We now indicate how (13) comes about. Let $0 \leqslant s < t$ and define

$$\varphi(z) = \varphi(z, s, t) = f_t^{-1} \circ f_s(z) = e^{s-t}z + \cdots.$$

This φ is a holomorphic map from U into U, but not onto U, such that 0 is carried to 0. Hence by the Schwarz lemma,

$$|\varphi(z, s, t)| < |z| = |\varphi(z, s, s)|$$

for all $z \neq 0$ ("inward flow on the unit disc"). Let us assume that $\partial\varphi/\partial t$ exists (and is analytic in z). Then the angle between the vector $\partial\varphi/\partial t$ for $t = s$ and the vector $-z$ must be bounded by $\frac{1}{2}\pi$. It follows that

$$\left.\frac{\partial \varphi}{\partial t}\right|_{t=s} = -zp(z, s) \text{ with } \operatorname{Re} p(z, s) > 0 \tag{14}$$

and $p(z, s)$ analytic in z, $p(0, s) = 1$.

From the definition of φ,

$$f_t \circ \varphi(z, s, t) = f_s(z).$$

Differentiating with respect to t and then setting $t = s$, we obtain

$$\frac{\partial f_t}{\partial t} + \frac{\partial f_t}{\partial z}\frac{\partial \varphi}{\partial t} = 0 \quad \text{for } t = s. \tag{15}$$

Combination of (15) and (14) gives Löwner's equation for $t = s$.

The assumption that φ is a nice function of its arguments is no real restriction, since we may assume that our domains D_t depend analytically on t. However, the Löwner differential equation holds for arbitrary Löwner chains, cf. Pommerenke [**14**] Chapter 6. A crucial observation (which makes use of the distortion formula (4)) is that $f(z, t)$ is Lipschitzian in t; equation (13a) will hold for almost all t. Conversely, every partial differential equation (13) determines a Löwner chain of conformal maps.

(iv) *Logarithmic coefficients for* $f(z, t)/e^t$. It is natural to introduce the expansions

$$\log\frac{f(z, t)}{e^t z} = \sum_1^\infty c_k(t) z^k. \tag{16}$$

Since $f(z, t)/e^t$ is in S, cf. (12b), we know from Section 2 that there exist constants A_k, for example $A_k = ek^2$, such that $|a_k(t)| \leqslant A_k$ for all t. Hence by recursion, cf. equation (10), there will be constants C_k such that

$$(17) \qquad |c_k(t)| \leqslant C_k \quad \text{for all } t.$$

We may differentiate relation (16) with respect to t and with respect to z. We substitute the results in the Löwner equation (13a), setting

$$(18) \qquad p(z,t) = 1 + 2\sum_1^\infty d_k(t) z^k.$$

Equating the coefficients of like powers of z, we thus obtain the system of differential equations

$$(19) \qquad c_k'(t) = 2d_k(t) + kc_k(t) + 2\sum_1^{k-1} jc_j(t) d_{k-j}(t), \quad k = 1,2,\dots .$$

(v) *The auxiliary functional* Ω. We now take *n fixed*. With an eye to the L-M conjecture (9) and following de Branges' ideas, cf. also [**8**], we introduce the auxiliary functional

$$(20) \qquad \Omega(t) = \Omega_n(t) = \sum_1^{n-1} \left\{ k|c_k(t)|^2 - \frac{4}{k} \right\} \sigma_k(t),$$

where the weight functions $\sigma_k(t)$ are to be chosen in a suitable manner. What properties besides some smoothness do we want the $\sigma_k(t)$ to have?

It is desired that the relation $\Omega(0) \leqslant 0$ be the conjectured L-M inequality (9). Noting that $c_k(0) = c_k$, we therefore impose the initial conditions

$$(21) \qquad \sigma_k(0) = n - k, \quad k = 1,\dots,n-1.$$

Clearly the inequality $\Omega(0) \leqslant 0$ would follow if $\Omega(t)$ were a non-decreasing function of t which vanishes at $t = +\infty$, that is, if

$$(22) \qquad \Omega'(t) \geqslant 0 \quad \text{for } 0 \leqslant t < \infty,$$

while $\Omega(t) \to 0$ as $t \to \infty$. Because of the boundedness of every $c_k(t)$, cf. (17), the last condition will be satisfied if

$$(23) \qquad \lim_{t\to\infty} \sigma_k(t) = 0, \quad k = 1,\dots,n-1.$$

Do there really exist functions $\sigma_k(t)$ satisfying (21) and (23) such that at the same time $\Omega'(t) \geqslant 0$?

(vi) *Differential equation conditions on the* σ_k *in order to make* Ω' *manageable*. We calculate $\Omega'(t)$ using the differential equations (19) for the $c_k(t)$. The resulting expression is quite complicated. However, after some experimentation it is seen to simplify if we impose de Branges' conditions

$$(24) \qquad \sigma_k - \sigma_{k+1} = -\left(\frac{\sigma_k'}{k} + \frac{\sigma_{k+1}'}{k+1} \right), \quad k = 1,\dots,n-1; \quad \sigma_n \equiv 0,$$

where the variable t has been suppressed. The result of the calculation may then be written in the form

$$(25) \qquad \Omega' = -\sum_1^{n-1} Q_k(c,d)\sigma_k',$$

where the Q_k are nonnegative functions of the $c_k(t)$ and the $d_k(t)$.

Since it is of importance for the case of equality in the L-M conjecture, we indicate the precise form of $Q_k(c,d)$. Using the Herglotz representation for holomorphic functions on the unit disc with positive real part, we have

$$p(z,t) = \int_{-\pi}^{\pi} \frac{e^{i\theta} + z}{e^{i\theta} - z}\, d\mu_t(\theta),$$

where μ_t is a positive Borel measure of total mass equal to $p(0,t) = 1$. Thus by (18), the coefficients of $p(z,t)$ have the form

$$d_k(t) = \int_{-\pi}^{\pi} e^{-ik\theta}\, d\mu_t(\theta).$$

Introducing the sums

$$s_k = \sum_{j=1}^{k} jc_j(t)\, e^{ij\theta}, \quad s_0 = 0 \tag{26}$$

we may write $kc_k(t) = (s_k - s_{k-1})e^{-ik\theta}$ and by (19)

$$c_k' = \int_{-\pi}^{\pi} (2 + s_{k-1} + s_k)\, e^{-ik\theta}\, d\mu_t(\theta).$$

Ω' may now be written as an integral relative to μ_t; the integrand is a sum involving the s_k, σ_k and σ_k'. After a summation by parts and from the differential equation (24), the result is (25) with

$$Q_k(c,d) = \frac{1}{k}\int_{-\pi}^{\pi} |2 + s_{k-1} + s_k|^2 d\mu_t. \tag{27}$$

(vii) *Explicit form of the* σ_k'. By the formula for Ω' (25), we would have its positivity (22), if we could *guarantee* that

$$\sigma_k' \leqslant 0, \quad k = 1, \ldots, n-1. \tag{28}$$

Observe, however, that first σ_{n-1} and next $\sigma_{n-2}, \ldots, \sigma_1$ are *completely determined* by the system of differential equations (24) and the initial conditions (21)! Could it be true that for the solutions, the additional conditions (23) and (28) are miraculously satisfied?

EXAMPLES (cf. the examples in Section 2). For $n = 2$, one has $\sigma_2 \equiv 0$ and hence $\sigma_1 = e^{-t}$. Thus $\sigma_1' \leqslant 0$. It follows that the L-M inequality holds for $n = 2$ and thus also the Bieberbach inequality $|a_2| \leqslant 2$.

For $n = 3$, one has $\sigma_3 \equiv 0$ and next

$$\sigma_2 = e^{-2t}, \quad \sigma_1 = 4e^{-t} - 2e^{-2t}.$$

Again $\sigma_k' \leqslant 0$, thus proving the L-M inequality for $n = 3$ and hence Löwner's inequality $|a_3| \leqslant 3$.

Of course, de Branges went on. For general n he found a solution of his system of differential equations and initial conditions which may be written as

$$\sigma_k(t) = k \sum_{\nu=0}^{n-k-1} (-1)^{\nu} \frac{(2k+\nu+1)_{\nu}(2k+2\nu+2)_{n-k-1-\nu}}{(k+\nu)\,\nu!(n-k-1-\nu)!} e^{-\nu t - kt},$$

$k = 1, \ldots, n-1$. Here

$$(a)_{\nu} = a(a+1)\cdots(a+\nu-1) \quad \text{for } \nu \geqslant 1, (a)_0 = 1.$$

It is clear that the functions $\sigma_k(t)$ will vanish at infinity (condition (23)). However, what about the negativity condition $\sigma_k' \leqslant 0$ in (28)? In other words, could it be true that the sums

$$-\frac{\sigma_k'}{k} e^{kt} = \sum_{\nu=0}^{n-k-1} (-1)^{\nu} \frac{(2k+\nu+1)_{\nu}(2k+2\nu+2)_{n-k-1-\nu}}{\nu!(n-k-1-\nu)!} e^{-\nu t} \tag{29}$$

are nonnegative for $k = 1, \ldots, n-1$ and all $n \geqslant 2$?

(viii) *Completion of the proof of the L-M conjecture*. For relatively small n, de Branges could verify immediately that the sums (29) are positive on $(0, \infty)$. But what about larger values of n? At this stage de Branges went to his numerical colleague Gautschi at Purdue University for help. He told Gautschi that he had a way of proving the Bieberbach conjecture, but needed to establish

certain inequalities involving hypergeometric functions. Would Gautschi be willing to check as many of these inequalities as possible on the computer? Gautschi wrote a suitable program with a feeling that he might soon hit a value of n for which the consistent positivity of expressions related to (29) would come to an end. Much to his surprise, however, he discovered that the crucial expressions were positive for all values of n which he tried: $2 \leqslant n \leqslant 30$. Thus at this time, assuming that the theoretical work was correct, de Branges and the computer had verified the Bieberbach conjecture for all n up to 30!

How to continue? Gautschi had the idea to call Askey at the University of Wisconsin, the world's expert on special functions. At first Askey was incredulous that the supposed positivity of sums such as those in (29) would prove the Bieberbach conjecture. However, he realized very soon that those sums were essentially generalized hypergeometric functions of a very special kind which are known to be positive:

$$-\frac{\sigma_k'}{k}e^{kt} = \binom{n+k}{n-k-1}{}_3F_2\left(\begin{matrix}-n+k+1, k+\frac{1}{2}, n+k+1\\ k+\frac{3}{2}, 2k+1\end{matrix}\,\middle|\, e^{-t}\right). \tag{30}$$

Here

$${}_3F_2\left(\begin{matrix}a, b, c\\ d, e\end{matrix}\,\middle|\, x\right) = \sum_{\nu=0}^{\infty}\frac{(a)_\nu(b)_\nu(c)_\nu}{(d)_\nu(e)_\nu}\frac{x^\nu}{\nu!};$$

for the special values of a through e in (30), the positivity of ${}_3F_2$ followed from a joint result of Askey and Gasper [**1**].

Thus de Branges' proof of the L-M conjecture was complete, thanks to a known result on special functions.

(ix) *The case of equality*. We now start with an arbitrary function $f \in S$ and an associated Löwner chain. It is easy to see that equality holds in the L-M inequality (9) for f and given $n \geqslant 2$ if and only if $\Omega' \equiv 0$. Since $\sigma_k' < 0$ on $(0, \infty)$ for $1 \leqslant k \leqslant n-1$, the latter condition requires that $Q_k(c, d) \equiv 0$ for those values of k, cf. (25). In particular, the condition $Q_1(c, d) \equiv 0$ is necessary for equality. By the representation (27) with positive μ_t this condition implies

$$2 + s_1 = 2 + c_1(t)e^{i\theta} = 0 \text{ a.e. } [\mu_t].$$

Thus the absolutely continuous part of μ_t must be zero, and in fact, μ_t must be a point mass 1 at some point θ_t. It follows that $|c_1(t)| \equiv 2$ and in particular $|c_1| = 2$, hence $|a_2| = 2$ so that f must be a Koebe function. For a Koebe function, one indeed has $\Omega(0) = 0$ for every n.

For the case of equality in the "Bieberbach inequality" (3), one may now use the remark at the end of Section 2.

5. Final remarks. De Branges was born in Paris in 1932. He studied in the U.S. and has been at Purdue University since 1963. In his mathematical career, he has tackled a number of difficult problems. Early in 1984 he completed a manuscript of 385 pages for a new edition of his book "Square summable power series". This manuscript culminated in a proof of the Bieberbach conjecture. With the manuscript, de Branges departed for Leningrad in April 1984 for a scheduled exchange visit. As he tells it, he was disappointed that the U.S. mathematicians to whom he had sent his manuscript had not yet been able to verify his long proof. In Leningrad, de Branges presented his work to the members of the seminars in functional analysis and geometric function theory. In a large number of sessions, the proof was verified and some inessential errors corrected. Finally, through hard work under de Branges' direction, a relatively short proof of the Lebedev-Milin conjecture was distilled from the original manuscript.

Upon his return from Leningrad, de Branges lectured on his proof at a number of universities, among them the Free University at Amsterdam. An early report on the proof in Russian was widely circulated. It reached FitzGerald and Pommerenke at La Jolla in July. They restated the

proof in their own words, as mathematicians do when they try to understand new material. They also treated the case of equality in the Bieberbach conjecture [**8**], as would others. In the mean time, de Branges produced a more sophisticated write-up of his proof which includes the case of equality in a very natural way [**4**]. Comments on the exciting events have been written up by FitzGerald [**7**] and by Gautschi [**9**], among others. (Added in proof: see also Pommerenke [**15**].)

References

1. R. Askey and G. Gasper, Positive Jacobi polynomial sums II, Amer. J. Math., 98 (1976) 709–737.

2. L. Bieberbach, Über die Koeffizienten derjenigen Potenzreihen, welche eine schlichte Abbildung des Einheitskreises vermitteln, S.-B. Preuss. Akad. Wiss., 1916, pp. 940–955.

3. L. de Branges, A proof of the Bieberbach conjecture, preprint E-5-84, Steklov Math. Inst., Leningrad Branch, 1984, pp. 1–21.

4. ______, A proof of the Bieberbach conjecture, Acta Math., 154 (1985) 137–152.

5. ______, Square summable power series, 2nd ed., to appear.

6. P. L. Duren, Univalent Functions, Springer, Berlin, 1983.

7. C. H. FitzGerald, The Bieberbach conjecture: retrospective. Notices, Amer. Math. Soc., 32 (1985) 2–6.

8. C. H. FitzGerald and C. Pommerenke, The de Branges theorem on univalent functions, Trans. Amer. Math. Soc., 290 (1985) 683–690.

9. W. Gautschi, Reminiscences of my involvement in de Branges' proof of the Bieberbach conjecture, to appear.

10. G. M. Goluzin, Geometric theory of functions of a complex variable. English transl., Amer. Math. Soc., Providence, R.I., 1969.

11. A. W. Goodman, Univalent Functions, vols. I, II, Mariner, Tampa, FL, 1983.

12. K. Löwner, Untersuchungen über schlichte konforme Abbildungen des Einheitskreises, I, Math. Ann., 89 (1923) 103–121.

13. I. M. Milin, Univalent functions and orthonormal systems. English trans., Amer. Math. Soc., Providence, RI, 1977.

14. C. Pommerenke, Univalent Functions, Vandenhoeck and Ruprecht, Göttingen, 1975.

15. ______, The Bieberbach conjecture, Math. Intelligencer, 7 (1985) 23–25, 32.

NON-SEXIST SOLUTION OF THE MÉNAGE PROBLEM

KENNETH P. BOGART
Department of Mathematics, Dartmouth College, Hanover, NH 03755

PETER G. DOYLE
AT & T Bell Labs, Murray Hill, NJ 07974

1. The ménage problem. The *ménage problem* (problème des ménages) asks for the number M_n of ways of seating n man-woman couples at a circular table, with men and women

---

Kenneth Bogart received his B.S. in Mathematics from Marietta College in 1965 and his Ph.D. in Mathematics from CalTech in 1968. He has been teaching at Dartmouth College, where he is a professor of Mathematics and Computer Science, ever since. His research interests are in the applications of algebra to combinatorics, and the applications of combinatorics, especially to Computer Science. Much of his research has involved the theory or applications of ordered sets. His textbook, *Introductory Combinatorics*, was published in 1983 by Pitman Publishing Company. He is currently writing a textbook on discrete mathematics.

Peter Doyle got his Ph.D. in mathematics from Dartmouth College in 1982. While he was at Dartmouth, he and Laurie Snell wrote a book called *Random Walks and Electric Networks* which was published by the MAA in 1984 in the Carus Mathematical Monographs series. After getting his degree he worked as a systems programmer for a year, and spent a year at the Institute for Mathematics and its Applications in Minneapolis. He now works in the Mathematical Sciences Research Center at AT & T Bell Labs in Murray Hill, NJ.

alternating, so that no one sits next to his or her partner. This famous problem was initially posed by Lucas [**8**] in 1891, though an equivalent problem had been raised earlier by Tait [**12**] in connection with his work on knot theory (see Kaplansky and Riordan [**6**]). This problem has been discussed by numerous authors (see the references listed in [**6**]), and many solutions have been found. Most of these solutions tell how to compute M_n using recurrence relations or generating functions, as opposed to giving an explicit formula. The first explicit formula for M_n was published by Touchard [**13**] in 1934, though he did not give a proof. Finally, in 1943, Kaplansky [**5**] gave a proof of Touchard's formula. Kaplansky's derivation was simple but not quite straightforward, and the problem is still generally regarded to be tricky.

We will present a completely straightforward derivation of Touchard's formula. Like Kaplansky's, our solution is based on the principle of inclusion and exclusion (see Ryser [**11**] and Riordan [**9**]). What distinguishes our approach is that we do not seat the ladies (or gentlemen) first.

2. Solution to the relaxed ménage problem. We begin with an apparently simpler problem, called the *relaxed ménage problem*, which asks for the number m_n of ways of seating n couples around a circular table so that no one sits next to his or her partner. This is nearly the same as the ménage problem, only now we have relaxed the requirement that men and women alternate.

To determine m_n, we begin with the set S of all $(2n)!$ ways of seating the $2n$ individuals around the table, and use inclusion-exclusion on the set of couples who end up sitting together. Let us call the elements of S seatings, and let us denote by w_k the number of seatings under which some specified set of k couples (and possibly some other couples) end up sitting together. Clearly, w_k does not depend on the particular set of k couples we choose, and so, by the principle of inclusion and exclusion, we have

$$m_n = \sum_{k=0}^{n} (-1)^k \cdot \binom{n}{k} \cdot w_k.$$

To finish the enumeration, we must compute w_k. Assume $n > 1$. Let d_k denote the number of ways of placing k non-overlapping unlabeled dominos on $2n$ vertices arranged in a circle. (See Fig. 1.) Then

$$w_k = d_k \cdot k! \cdot 2^k \cdot (2n - 2k)!.$$

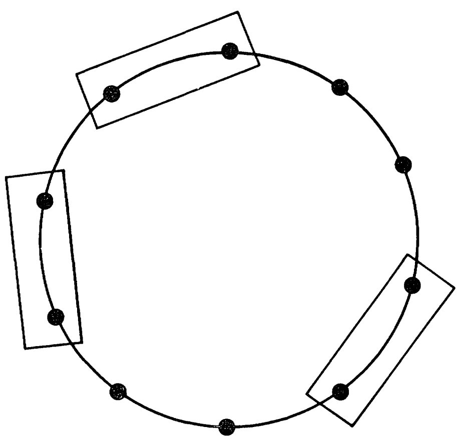


FIG. 1. Non-overlapping dominos.

(Decide where the k couples go, and which couple goes where, and which partner takes which seat, and where the $2n - 2k$ individuals go.) So now we have only to compute the d_k's. This is a routine combinatorial problem. The answer is

$$d_k = \frac{2n}{2n-k} \cdot \binom{2n-k}{k}$$

(see Ryser [**11**], pp. 33–34, or Exercise 1 below). This yields

$$w_k = 2n \cdot (2n-k-1)! \cdot 2^k.$$

Plugging this expression for w_k into the formula for m_n above, we get

$$m_n = \sum_{k=0}^{n} (-1)^k \cdot \binom{n}{k} \cdot 2n \cdot (2n-k-1)! \cdot 2^k.$$

By symmetry, we know that m_n must be divisible by $2^n \cdot n!$. Pulling this factor out in front, we can write

$$m_n = 2^n \cdot n! \cdot \sum_{k=0}^{n} (-1)^k \cdot \frac{2n}{2n-k} \cdot \binom{2n-k}{k} \cdot (1 \cdot 3 \cdot 5 \cdots (2n-2k-1)).$$

The first few values of m_n are shown in Table 1.

TABLE 1. Relaxed ménage numbers.

n	m_n	$m_n/(2^n n!)$	$m_n/(2n)!$
2	8	1	0.333333...
3	192	4	0.266666...
4	11904	31	0.295238...
5	1125120	293	0.310052...
6	153262080	3326	0.319961...
7	28507207680	44189	0.326998...
8	6951513784320	673471	0.332246...
9	2153151603671040	11588884	0.336305...
10	826060810479206400	222304897	0.339537...

3. Solution to the ménage problem. For the ménage problem, we proceed just as before, only now we restrict the set S of seatings to those where men and women alternate. The number of these seatings is $2(n!)^2$: two ways to choose which seats are for men and which for women; $n!$ ways to seat the men in the men's seats; $n!$ ways to seat the women in the women's seats. Just as before, we have

$$M_n = \sum_{k=0}^{n} (-1)^k \cdot \binom{n}{k} \cdot W_k,$$

where W_k denotes the number of alternating seatings under which a specified set of k couples all end up sitting together. This time we have

$$W_k = 2 \cdot d_k \cdot k! \cdot (n-k)!^2.$$

(Decide which are men's seats and which women's, where the k couples go, which couple goes where, and where the $n-k$ men and $n-k$ women go.) Plugging in for d_k yields

$$W_k = 2 \cdot 2n \cdot (2n-k-1)! \cdot \frac{(n-k)!^2}{(2n-2k)!}.$$

Plugging this expression for W_k into the formula for M_n above, we get

$$M_n = \sum_{k=0}^{n} (-1)^k \cdot \binom{n}{k} \cdot 2 \cdot 2n \cdot (2n-k-1)! \cdot \frac{(n-k)!^2}{(2n-2k)!}.$$

By symmetry, we know that M_n must be divisible by $2 \cdot n!$. Pulling this factor out in front, we can write

$$M_n = 2 \cdot n! \cdot \sum_{k=0}^{n} (-1)^k \cdot \frac{2n}{2n-k} \cdot \binom{2n-k}{k} \cdot (n-k)!.$$

The first few values of M_n are shown in Table 2.

TABLE 2. Ménage numbers.

n	M_n	$M_n/(2n!)$	$M_n/(2n!^2)$
2	0	0	0.0
3	12	1	0.166666...
4	96	2	0.083333...
5	3120	13	0.108333...
6	115200	80	0.111111...
7	5836320	579	0.114880...
8	382072320	4738	0.117509...
9	31488549120	43387	0.119562...
10	3191834419200	439792	0.121194...

4. Comparison with Kaplansky's solution. The solution that we have just given is completely straightforward and elementary, yet we have said that the ménage problem is still generally regarded to be tricky. How can this be? The answer can be given in two words: "Ladies first." It apparently never occurred to anyone who looked at the problem not to seat the ladies first (or in a few cases, the gentlemen). Thus Kaplansky and Riordan [**6**]: "We may begin by fixing the position of husbands and wives, say wives for courtesy's sake."

Seating the ladies first "reduces" the ménage problem to a problem of permutations with restricted position. Unfortunately, this new problem is more difficult than the problem we began with, as we may judge from the cleverness of Kaplansky's solution [**5**]:

> We now restate the *problème des ménages* in the usual fashion by observing that the answer is $2n!u_n$, where u_n is the number of permutations of $1,\dots,n$ which do not satisfy any of the following $2n$ conditions: 1 is 1st or 2nd, 2 is 2nd or 3rd, ..., n is nth or 1st. Now let us select a subset of k conditions from the above $2n$ and inquire how many permutations of $1,\dots,n$ there are which satisfy all k; the answer is $(n-k)!$ or 0 according as the k conditions are compatible or not. If we further denote by v_k the number of ways of selecting k compatible conditions from the $2n$, we have, by the familiar argument of inclusion and exclusion, $u_n = \Sigma(-1)^k v_k (n-k)!$. It remains to evaluate v_k, for which purpose we note that the $2n$ conditions, when arrayed in a circle, have the property that only consecutive ones are not compatible.... .

Of course $v_k = d_k$, so we see how, by choosing to view the constraints as arrayed in a circle, Kaplansky has gotten back on the track of the straightforward solution. We can only admire Kaplansky's cleverness in rediscovering the circle, and regret the tradition of seating the ladies first that made such cleverness necessary.

5. Conclusion. It appears that it was only the tradition of seating the ladies first that made the ménage problem seem in any way difficult. We may speculate that, were it not for this tradition, it would not have taken half a century to discover Touchard's formula for M_n. Of all the ways in which sexism has held back the advance of mathematics, this may well be the most peculiar. (But see Exercise 2.)

6. Exercises. We list here, in the guise of exercises, some questions that you may want to explore with the help of the references listed.

1. Show how to "derive" the formula for d_k simply by writing down the answer, without using recurrence relations or generating functions or what have you. (Hint: Try this first for the formula for w_k.)
2. Was it really sexism that made the ménage problem appear difficult? (See Kaplansky and

Riordan [6], and the references listed there.)

3. Solve the analog of the ménage problem for the situation depicted in Figure 2. (No one is allowed to sit next to or across from his or her partner.)

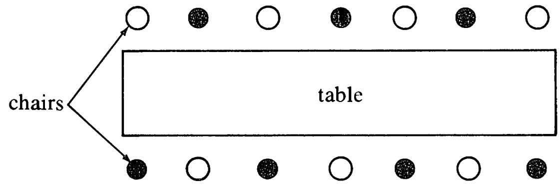


FIG. 2. Real-world ménage problem.

4. Formulate the analog of the ménage problem for an arbitrary graph G, and show that it leads to a domino problem on G. Show that by seating the ladies or gentlemen first, and following Kaplansky's lead, we arrive at a problem of how to place rooks on a chessboard. (See Riordan [9], Chap. 7.) Show that the domino problem and the rook problem are equivalent. Look into the relationship of the domino problem to the Ising model of statistical mechanics. (See Fisher [3], Kasteleyn [7].)

5. What problem was Tait [12] really interested in? Did Gilbert [4] solve it? Show that Gilbert could have used a simple Möbius inversion argument instead of Pólya's theorem. What kinds of problems require the full force of Pólya's theorem?

6. What does it mean to "solve" a combinatorial problem like the ménage problem? Is a closed-form solution better than a recurrence? What if what we really want is to generate configurations, rather than just count them? (See Wilf [14].)

7. Why did Tait not pursue the ménage problem? What do knots have to do with atomic spectra? What was it like to live in Nebraska in the 1880's? (See Conway [2].)

8. The relaxed ménage problem can be further generalized as follows: Given two graphs G_1 and G_2 with the same number of vertices, find the number of one-to-one mappings of the vertices of G_1 onto the vertices of G_2 such that no pair of vertices that are adjacent in G_1 get sent to vertices that are adjacent in G_2. Show that the dinner table problem (see Aspvall and Liang [1], Robbins [10]) can be phrased in these terms, and give a solution using inclusion-exclusion. Formulate and solve an "unrelaxed" version of this problem. Show that the ménage problem can be phrased in these terms, and discuss how useful this reformulation is. Do the same for the problem of enumerating Latin rectangles (see Ryser [11]).

References

1. B. Aspvall and F. M. Liang, The dinner table problem, Report STAN-CS-80-8222, Computer Science Department, Stanford University, Stanford, CA, 1980.

2. J. H. Conway, An enumeration of knots and links, and some of their algebraic properties, pp. 329–358 in Computational Problems in Abstract Algebra, ed. J. Leech, Pergamon, Oxford, 1970.

3. M. E. Fisher, Statistical mechanics of dimers on a plane lattice, Phys. Rev., 124 (1961) 1664–1672.

4. E. N. Gilbert, Knots and classes of ménage permutations, Scripta Math., 22 (1956) 228–233.

5. I. Kaplansky, Solution of the problème des ménages, Bull. Amer. Math. Soc., 49 (1943) 784–785.

6. I. Kaplansky and J. Riordan, The problème des ménages, Scripta Math., 12 (1946) 113–124.

7. P. W. Kasteleyn, Dimer Statistics and Phase Transitions, J. Math. Phys., 4 (1963) 287–293.

8. E. Lucas, Théorie des nombres, Gauthier-Villars, Paris, 1891.

9. J. Riordan, An Introduction to Combinatorial Analysis, Wiley, New York, 1958.

10. D. Robbins, The probability that neighbors remain neighbors after random rearrangements, this MONTHLY, 87 (1980) 122–124.

11. H. J. Ryser, Combinatorial Mathematics, Carus Mathematical Monograph No. 14, Mathematical Association of America, Washington, DC, 1963.

12. P. G. Tait, On knots, I, II, III, pp. 273–347 in Scientific Papers, Cambridge Univ. Press, Cambridge, 1898.

13. J. Touchard, Sur un problème des permutations, C. R. Acad. Sciences Paris, 198 (1934) 631–633

14. H. Wilf, What is an answer?, this MONTHLY, 89 (1982) 289–292.

Should her subject be called geometric analysis or analytic geometry? (See p. 530.)

```
            { To modify this program for the case of the half
            deck, j = nk/2, let the above loop run from I:=0 TO
            TRUNC(K*N[TRIALS]/2)-1. }
            E:=E/(K*N[TRIALS]);
            WRITELN ('THE EXPECTED # OF HITS FOR A DECK OF
            SIZE');
            WRITELN (N[TRIALS]:3,'*4':2,' IS ':4,E:5:2);
            WRITELN;
      END;
END.
```

References

L. Bloom, Language Development: Form and Function in Emerging Grammars, M.I.T. Press, Cambridge, MA, 1970.

R. Brown, A First Language: The Early Stages, Harvard University Press, Cambridge, MA, 1973.

B. T. Gardner and R. A. Gardner, Two-way communication with an infant chimpanzee, in A. Schrier and F. Stollnitz (eds.), Behavior of Nonhuman Primates (Vol. 4 (1971) 117–184), Academic Press, New York.

R. A. Gardner and B. T. Gardner, Comparative psychology and language acquisition, Ann. of the New York Academy of Sciences, 309 (1978) 37–76.

R. A. Gardner and B. T. Gardner, A vocabulary test for chimpanzees, J. of Comparative Psychology, 98 (1984) 381–404.

K. J. Hayes and C. Hayes, The intellectual development of a home-raised chimpanzee, Proceedings of the American Philosophical Society, 95 (1951) 105–109.

H. M. Jenkins, Measurement of stimulus control during discriminative operant conditioning, Psychological Bulletin, 64 (1965) 365–376.

W. N. Kellogg and L. A. Kellogg, The Ape and the Child, Hafner, New York, 1933.

R. C. Read, Card-guessing with information—a problem in probability, this MONTHLY, 69 (1962) 506–511.

J. Umiker-Sebeok and T. A. Sebeok, Questioning apes, in T. A. Sebeok and J. Umiker-Sebeok, (eds.) Speaking of apes (pp. 1–59), Plenum, New York, 1980.

---

172. MISCELLANEA

In a world in which the price of calculation continues to decrease rapidly, but the price of theorem proving continues to hold steady or increase, elementary economics indicates that we ought to spend a larger and larger fraction of our time on calculation.

Source: John W. Tukey, *The American Statistician*, 40 (1986) p. 74.

Editorial comment: !

ANSWER TO PHOTO ON PAGE 519

Karen Uhlenbeck.

CHANCE EXPECTANCY WITH TRIAL-BY-TRIAL FEEDBACK AND RANDOM SAMPLING WITHOUT REPLACEMENT*

JAMES C. PATTERSON
Department of Mathematics, Eastern Kentucky University, Richmond, KY 40475

BEATRIX T. GARDNER AND R. ALLEN GARDNER
Department of Psychology, University of Nevada Reno, Reno, NV 89557-0062

Most of the easily accessible statistical models for estimating chance performance assume random sampling with replacement, but the objectives of psychological experiments commonly require sampling without replacement. It is often desirable, for example, to administer a repeated series of measurements to the same individual while balancing conditions, types of stimulus event, or types of response alternative. Since both human and nonhuman subjects can detect the nonrandom effects of sampling without replacement (Jenkins, 1965), a precise estimate of chance performance must take this into account.

Consider the following experimental example generated by a well-known and hotly-disputed theoretical question. Gardner and Gardner (1978) reared five chimpanzees, Washoe, Moja, Pili, Tatu, and Dar, under conditions that closely simulated the rearing conditions of human infants. Where earlier experiments in cross-fostering such as Kellogg and Kellogg (1933) and Hayes and Hayes (1951) used speech, Gardner and Gardner used American Sign Language (ASL), the gestural language of the deaf in North America. Naturalistic observations of the young chimpanzees yielded results that closely approximate the results of similar observations of very young human children.

While other developmental similarities were accepted without dispute, the evidence for verbal skills generated heated controversy (e.g., Umiker-Sebeok & Sebeok, 1980). One of the more pertinent arguments that have been raised concerns the validity of adventitious, naturalistic observations, particularly in the case of verbal behavior. Under the rich, informal conditions of cross-fostering, observers may read more meaning than is actually present in the signs. The fact that the remarkable linguistic skills that have been attributed to very young human children are based on equally informal observations (c.f. Bloom, 1970; Brown, 1973) only serves to underscore the need for more stringent evidence.

Early in Project Washoe, Gardner and Gardner (1971, pp. 158–161) began a series of vocabulary tests designed to demonstrate that cross-fostered chimpanzees could communicate to a human observer under conditions in which the only information available to the observer was the sign language of the chimpanzees. To accomplish this, exemplars of nameable objects were photographed on 35-mm slides which were back-projected on a screen that could be seen by the chimpanzee subject, but could not be seen by the observers. A second objective was to demonstrate that independent observers agreed with each other. To accomplish this, there were two observers. The first observer served as interlocutor in the testing room with the chimpanzee.

James Patterson's 1979 Ph.D. dissertation in homological algebra was supervised by Edgar Enochs at the University of Kentucky. Most of his time is now spent teaching undergraduate mathematics and computer operating systems, and enjoying his horses on a farm in eastern Kentucky. He is the nephew of Beatrix Gardner whose 1959 D.Phil. in Zoology is from Oxford University and Allen Gardner whose 1954 Ph.D. in Psychology is from Northwestern University. The Gardners have studied jumping spiders, stickleback fish, laboratory rats, cross-fostered chimpanzees, and human beings and are intensely interested in mathematical tools for the description and understanding of all aspects of animal behavior.

*Research supported by the Spencer Foundation and grant MH 39043 from the National Institute of Mental Health. The authors also wish to thank the Department of Mathematics at Eastern Kentucky University for providing the extensive computer time required for this work. Requests for reprints should be addressed to R. A. Gardner, Department of Psychology, University of Nevada, Reno, NV 89557-0062.

The second observer was stationed in a second room and observed the subject from behind one-way glass. The two observers gave independent readings; they could not see each other or compare readings in any way until after the testing session (see Gardner and Gardner, 1984, for details).

In the first description of the tests, Gardner and Gardner (1971) estimated chance performance as k/kn, where n is the number of vocabulary items on a test and k is the number of exemplars of each vocabulary item. As Umiker-Sebeok and Sebeok (1980, p. 40) point out, however, k/kn is an estimate that only considers the guessing of the chimpanzee. The observers were familiar with the list of vocabulary items and they knew that there were exactly four exemplars of each item on each test. Although the slides were shown in random order, the observers could have used their knowledge of past trials to improve their guesses during the course of a test.

The Sebeoks cite the familiar example of gamblers who can win at Black Jack by keeping track of the cards that have already been played. They also cite research on extrasensory perception (ESP) that attempted to demonstrate improvement in telepathic skill under conditions that permit trial-by-trial feedback. The ESP research lead to Read's (1962) mathematical analysis of card-guessing with feedback, which was in turn the starting point for the following more general analysis of the problem.

Mathematical analysis. Read (1962) considered a common experiment on ESP in which there is a special deck of cards containing k exemplars of n types of card and $k = n$. We assume that the deck of cards lies face down on a table and the percipient tries to guess the top card in the deck. After each guess the top card is turned face up and left on the table so that the percipient can use all of the information of all of the past trials in guessing the next top card. After r guesses the remainder of the deck contains a_1 cards of type 1, a_2 cards of type 2,..., and a_n cards of type n, so that

$$a_1 + a_2 + \cdots + a_n = n^2 - r.$$

Let $E(a_1, a_2, \ldots, a_n)$ denote the expected number of hits as the percipient proceeds through the remainder of the deck. If $a_1 = \max(a_1, a_2, \ldots, a_n)$, then type 1 is the best guess for the next trial. The probability that guessing type 1 will result in a hit is $a_1/(n^2 - r)$ and the expectation is $1 + E(a_1 - 1, a_2, \ldots, a_n)$. The probability that guessing type 1 will result in a miss (top card of type i rather than type 1) is $a_i/(n^2 - r)$ and the expectation is $E(a_1, a_2, \ldots, a_i - 1, \ldots, a_n)$.

We can express the expected number of hits after r guesses in terms of the expected number of hits after $r + 1$ guesses:

$$(1) \qquad E(a_1, a_2, \ldots, a_n) = \frac{a_1}{n^2 - r}(1 + E(a_1 - 1, a_2, \ldots, a_n)) + \frac{1}{n^2 - r}\left[\sum_{i=2}^{n} a_i E(a_1, a_2, \ldots, a_i - 1, \ldots, a_n)\right].$$

Proceeding from these assumptions, Read showed that the expected number of hits for the popular ESP deck of $k = n = 5$ cards is 8.65 for each run of 25 trials. This is, of course, appreciably better than $k/kn = 5$ hits expected without trial-by-trial feedback.

The application of Read (1962) to cases where $k \neq n$ is straightforward. The probability of a hit is $a_1/(kn - r)$ and the probability of a miss is $a_i/(kn - r)$. In general we can express the expected number of hits after r guesses as

$$(2) \quad E(a_1, a_2, \ldots, a_n) = \frac{1}{kn - r}\left[\max(a_1, a_2, \ldots, a_n) + \sum_{i=1}^{n} a_i E(a_1, a_2, \ldots, a_i - 1, \ldots, a_n)\right].$$

Direct application of (2) would require k^n recursive calls which presents formidable practical difficulties for all of the experimentally useful cases listed in Table 2. Fortunately, as Read observed, we can make use of the following symmetry:

$E(a_1, a_2, \ldots, a_n) = E(b_1, b_2, \ldots, b_n)$ if $\langle b_1, b_2, \ldots, b_n \rangle$ is a permutation of $\langle a_1, a_2, \ldots, a_n \rangle$.

If we let $E(a_1, a_2, \ldots, a_n)$ be represented by the monomial $X_1^{a_1} X_2^{a_2} \cdots X_n^{a_n}$, then the summation in (2) can be represented by applying the operator $\partial/\partial X_1 + \partial/\partial X_2 + \cdots + \partial/\partial X_n$ to the monomial $X_1^{a_1} X_2^{a_2} \cdots X_n^{a_n}$. Following Read we will refer to the use of symmetry to combine monomials as *condensation*.

Table 1 is a worked example for the case $k = 4$, $n = 2$.

Table 1

Worked Example for the Case of $\underline{k}$ = 4 Exemplars of $\underline{n}$ = 2 Types

P_0	$X_1^4 X_2^4$	
P_1	$\frac{1}{8}(8X_1^3X_2^4)$	$\frac{1}{2}$
P_2	$\frac{1}{8 \cdot 7}(24X_1^2X_2^4 + 32X_1^3X_2^3)$	$\frac{4}{7}$
P_3	$\frac{3!}{8!}(288X_1^2X_2^3 + 48X_1X_2^4)$	$\frac{4}{7}$
P_4	$\frac{4!}{8!}(48X_2^4 + 864X_1^2 + 768X_1X_2^3)$	$\frac{22}{35}$
P_5	$\frac{3!}{8!}(960X_2^3 + 5760X_1X_2^2)$	$\frac{22}{35}$
P_6	$\frac{2!}{8!}(8640X_2^2 + 11520X_1X_2)$	$\frac{5}{7}$
P_7	$\frac{1!}{8!}(40320X_2)$	$\frac{5}{7}$
P_8	$\frac{0!}{8!}(40320)$	1
	Total number of expected hits =	5.3

In each case in Table 1, the polynomial P_r together with the constant term on the right is calculated by applying (2) to P_{r-1} followed by condensation. Thus, in general

$$P_r = [(kn - r)!/(kn)!][\partial/\partial X_1 + \partial/\partial X_2 + \cdots + \partial/\partial X_n]^r X_1^k X_2^k \cdots X_n^k$$

followed by condensation.

Even with an algorithm to make optimum use of condensation and to avoid unnecessary duplication of effort, however, the recursive use of (2) presents formidable practical difficulties for most of the experimentally useful cases listed in Table 2. The minimum number of recursive calls would still be $\binom{n+k}{k}$ which is the number of integer lattice points in the nonnegative, n-dimensional orthant $\leqslant (k, k, \ldots, k)$.

To relieve ourselves of the burden of recursion, we must derive a closed form. Secondly, the expression must yield the expected number of hits for partial as well as complete runs through the deck because, as we shall explain presently, an important class of experimental cases involves partial runs. We begin by noting that it follows from (2) that the sum of the constants to the right and above P_r for $r = 1, 2, \ldots, kn$ equals the expected number of hits for a fractional run through the deck of r/kn. The expected number of hits for a run that ends 3/8 of the way through the deck of the example in Table 1 would equal $1/2 + 4/7 + 4/7$. Thus, both objectives can be accomplished by a general expression for the constant term to the right of P_r.

The first step is to find an expression for the coefficient of a typical term of P_r prior to condensation. Let p_i = the number of indeterminates taken to the ith power, where $i = 0, 1, 2, \ldots, k$. Then

$$p_0 + p_1 + p_2 + \cdots + p_k = n$$

$$p_1 + 2p_2 + \cdots + kp_k = kn - r.$$

For the indeterminate X_j to be to the ith power, it must have been acted on by $(\partial/\partial X_j)^{k-i}$. This would introduce a coefficient of $k!/i!$.

Applying the multinomial theorem together with the above observations, we see that the coefficient of a typical term would be

$$\frac{(kn-r)!}{(kn)!} \cdot \frac{r!}{[k!]^{p_0}[(k-1)!]^{p_1}[(k-2!)]^{p_2}\cdots}\left(\frac{k!}{0!}\right)^{p_0}\left(\frac{k!}{1!}\right)^{p_1}\left(\frac{k!}{2!}\right)^{p_2}\cdots\left(\frac{k!}{k!}\right)^{p_k}$$

or more simply

$$\frac{(kn-r)!r!}{(kn)!}\binom{k}{0}^{p_0}\binom{k}{1}^{p_1}\binom{k}{2}^{p_2}\cdots\binom{k}{k}^{p_k}.$$

Since this term is symmetric to $n!/p_0!p_1!\cdots p_k!$ terms, condensation yields a single term with the coefficient

$$\frac{(kn-r)!r!}{(kn)!} \cdot \frac{n!}{p_0!p_1!\cdots p_k!}\binom{k}{0}^{p_0}\binom{k}{1}^{p_1}\binom{k}{2}^{p_2}\cdots\binom{k}{k}^{p_k}.$$

The contribution of this term to the constant occurring to the right of P_{r+1} will be its coefficient times the highest power times $1/(kn-r)$, where "highest power" refers to the highest of the powers of the variables in the monomial. For a given power of β the terms for which this power is a maximum in P_r will contribute the following amount to the constant to the right of P_{r+1}:

$$\text{(3)} \qquad \frac{\beta(kn-r)!n!}{(kn-r)(kn)!}\sum\frac{r!}{p_0!p_1!\cdots p_\beta!}\binom{k}{0}^{p_0}\binom{k}{1}^{p_1}\cdots\binom{k}{\beta}^{p_\beta},$$

where the summation is for

$$p_0 + p_1 + \cdots + p_\beta = n, \quad p_1 + p_2 + \cdots + \beta p_\beta = kn - r, \quad p_\beta > 0.$$

This summation is the coefficient of t^{kn-r} in

$$\left[\binom{k}{0} + \binom{k}{1}t + \cdots + \binom{k}{\beta}t^\beta\right]^n - \left[\binom{k}{0} + \binom{k}{1}t + \cdots + \binom{k}{\beta-1}t^{\beta-1}\right]^n,$$

where the second part of the expression subtracts out the contribution for $p_\beta = 0$.

If

$$f_\beta(t) = \left[\binom{k}{0} + \binom{k}{1}t + \cdots + \binom{k}{\beta}t^\beta\right]^n,$$

then (3) can be rewritten as

$$\frac{\beta}{kn\binom{kn-1}{r}} \text{ times the coefficient of } t^{kn-r} \text{ in } f_\beta(t) - f_{\beta-1}(t).$$

Summing over all the values of β, we find that the constant to the right of P_{r+1} is

$$\frac{A_{kn-r}}{kn\binom{kn-1}{r}},$$

where A_{kn-r} represents the coefficient of t^{kn-r} in

$$k[f_k(t) - f_{k-1}(t)] + (k-1)[f_{k-1}(t) - f_{k-2}(t)] + \cdots + 2[f_2(t) - f_1(t)] + f_1(t) - f_0(t)$$

$$= kf_k(t) - \sum_{i=1}^{k-1} f_i(t).$$

Thus the expected number of hits for going through the portion of the deck equal to j cards ($j = 1, 2, \ldots, kn$) is

$$\sum_{r=0}^{j-1} \frac{A_{kn-r}}{kn\binom{kn-1}{r}}. \tag{4}$$

Applying (4) to the case where $k = 4$ exemplars of $n = 2$ types, we have

$$\begin{aligned} 4f_4(t) - \sum_{i=0}^{3} f_i(t) &= 4\left[\binom{4}{0} + \binom{4}{1}t + \binom{4}{2}t^2 + \binom{4}{3}t^3 + \binom{4}{4}t^4\right]^2 \\ &\quad - \left[\binom{4}{0} + \binom{4}{1}t + \binom{4}{2}t^2 + \binom{4}{3}t^3\right]^2 \\ &\quad - \left[\binom{4}{0} + \binom{4}{1}t + \binom{4}{2}t^2\right]^2 - \left[\binom{4}{0} + \binom{4}{1}t\right]^2 - \left[\binom{4}{0}\right]^2 \\ &= 8t + 40t^2 + 120t^3 + 176t^4 + 176t^5 + 96t^6 + 32t^7 + 4t^8. \end{aligned}$$

The expected number of hits for the complete deck ($j = 8$ trials) is

$$\begin{aligned} &\frac{1}{8}\left[\frac{4}{1} + \frac{32}{7} + \frac{96}{21} + \frac{176}{35} + \frac{176}{35} + \frac{120}{21} + \frac{40}{7} + \frac{8}{1}\right] \\ &= \frac{1}{2} + \frac{4}{7} + \frac{4}{7} + \frac{22}{35} + \frac{22}{35} + \frac{5}{7} + \frac{5}{7} + 1 \\ &= 5.3, \end{aligned}$$

as in Table 1.

Note that (4) is a practical procedure for a wide range of experimental cases in that it requires the summation of only kn terms, while at the same time it can be used to obtain the expected number of hits for any value of j corresponding to a partial run through the deck.

Application. An important class of experimental cases requires partial rather than complete runs through the deck. Gardner and Gardner (1984, pp. 391–393), for example, limited the participation of each observer to half of the trials of any single test in order to demonstrate that several different observers could read the signs of the cross-fostered chimpanzees. This would be the recommended experimental procedure in any case because the effect of sampling without replacement increases as we proceed through the deck. The last card is completely predictable,

the next to the last card can have at most two values, and so on. Consequently, restricting each observer to half of each test reduced the number of hits that should be expected by chance. This procedure is equivalent to reshuffling the deck halfway through each run. Frequent reshuffling is, of course, the device that gambling casinos use to defeat card-counting customers.

Formula (4) can be used to find the expected number of hits for any fraction of a deck by setting j equal to the number of cards in any fraction r/kn of the deck. Thus, in the worked example of Table 1, with $k = 4$ exemplars and $n = 2$ types, the expected number of hits for a run through one half of the deck is $1/2 + 4/7 + 4/7 + 22/35 = 2.3$ hits, which is less than half of the value of 5.3 hits for a run through the whole deck.

Table 2

Predicted and Obtained Hits as a Percent of *nk* Trials When *k* = 4 for Selected Values of *n*

	Full Deck		Half Deck			
n	Calc	Obt	Calc	Obt	O_1	O_2
2	66.6	66.3	56.8	56.0	na	na
4	42.1	42.3	30.9	31.4	na	na
8	25.5	25.6	16.4	16.1	na	na
16	15.0	15.0	8.5	8.4	85.9	87.5
21	12.1	12.1	6.5	6.3	78.6	79.8
25	10.6	10.5	5.5	5.4	84.0	84.8
27	9.9	10.0	5.1	5.1	83.3	81.3
32	8.7	8.7	4.3	4.3	71.9	71.1
34	8.3	8.3	4.0	4.1	80.1	78.7
35	8.1	8.0	3.9	3.9	54.5	53.8
50	6.0	6.0	2.8	2.8	na	na

Note: Calc = calculated from Formula (4), Obt = Obtained from computer simulations of 1,000 trials, O_1 and O_2 = hits as read by independent observers O_1 and O_2 in Gardner & Gardner (1984).

The values of n for the vocabulary tests reported in Gardner and Gardner (1984) ranged from 16 to 35 with k always equal to 4 and j always equal to half of the deck or $nk/2$. The average number of hits that would be expected by chance if an observer with perfect memory always made the best guess was calculated by Formula (4) for runs of full decks and runs of half decks. As a check on Formula (4), the performance of a simulated observer facing the same task was obtained by Monte Carlo simulation using averages of 1,000 simulated runs with the same values of n, k, and j. The results are expressed in Table 2 as percent hits and show that expected hits

calculated by Formula (4) agree closely with the expected hits obtained by simulation.

Three values of n that are lower than those used in Gardner and Gardner (1984) and one larger value are included in Table 2 to show the overall shape of the function. As n increases, the percent of hits expected by chance declines sharply at first, and then very gradually, but remains slightly above k/kn throughout the range of values in Table 2.

The computer program that generated the values in Table 2 using Formula (4) appears in the Appendix. This program was executed on a VAX 730 computer using a G-floating format with a range of $.56 \times 10^{-308}$ to $.9 \times 10^{308}$. It is straightforward and can be adapted for use on any computer equipped with a standard version of Pascal and equivalent range. With minor editing for other values of j, k, and n, this program can be used for a wide range of experimental cases. The computer program that generated the values in Table 2 for the simulated observer can be obtained from the authors on request.

The Monte Carlo simulation required much more computer time than the direct solution of Formula (4), of course. Thus, the simulation program required 65.46 min of CPU time to generate the values for full decks in Table 2 as opposed to 1.68 min for the direct solution program in the Appendix. For half decks the times were 43.98 min and 1.57 min.

The percent of hits scored by the chimpanzee subjects in the tests administered by Gardner and Gardner (1984) also appear in Table 2. While there is close agreement between the independent observers, the hits obtained by the chimpanzees are of an entirely different order of magnitude from chance expectancy. Mathematically sophisticated readers may have been able to estimate the size of these differences without recourse to the complete analysis presented here. Psychologists, however, cannot always count on the mathematical sophistication of all parties to a hotly-contested debate. In any event, whether critical deviations from chance are very small or very large, they are best supported by the most rigorous mathematical treatment available.

In this case, the psychological dispute led to a fruitful collaboration with a mathematician. We hope that the mathematical result will be generally instructive and that it will be of value for a wide range of experimental cases in which there is random sampling without replacement and trial-by-trial feedback to the subjects.

Appendix

Program for Computing Expected Number of Hits with Formula (4)

<u>Note</u>. This program was executed on a VAX 730 using a G-floating format for the data type, DOUBLE. Comments in { } are included before each function for descriptive purposes.

```
PROGRAM HITS (INPUT,OUTPUT);

CONST

    K=4; KN=200;

{ In all applications considered K=4.  The largest value of n

considered = 50, hence largest deck size = 4*50 = 200. }

TYPE

    POLY=ARRAY[0..KN] OF DOUBLE;
```

```
VAR
    E:DOUBLE;
    I,L,TRIALS:INTEGER;
    N:ARRAY[1..11] OF INTEGER;
    F:ARRAY[1..K] OF POLY;
FUNCTION FACT(Y:DOUBLE):DOUBLE;
{ This function returns the value Y! . }
VAR
    F:DOUBLE;
BEGIN
    IF Y<=1 THEN F:=1
    ELSE F:=FACT(Y-1)*Y;
    FACT:=F;
END;


FUNCTION BIN(A,B:INTEGER):DOUBLE;
{ This function returns the value A!/(B!*(A-B)!. }
VAR
    I:INTEGER;
    F,A1,B1:DOUBLE;
BEGIN
    A1:=A;B1:=B;F:=1.0;
    IF B>A-B THEN
         BEGIN
              FOR I:=0 TO A-B-1 DO
                   F:=F*(A-I);
              BIN:=F/FACT(A1-B1)
         END
```

```
    ELSE
        BEGIN
            FOR I:=0 TO B-1 DO
                F:=F*(A-I);
            BIN:=F/FACT(B1);
        END;
END;
PROCEDURE POWER (VAR G:POLY; N:INTEGER);
{ Takes the polynomial, G , to the Nth power. }
VAR
    R:ARRAY[0..K,0..KN] OF DOUBLE;
    I,P,L,SIZE:INTEGER;
    H:POLY;

BEGIN
    FOR I:=0 TO K DO
            FOR P:=0 TO KN DO R[I,P]:=0;
    FOR I:=0 TO KN DO H[I]:=G[I];
    SIZE:=K;
    FOR L:=2 TO N DO
        BEGIN
            FOR I:=0 TO K DO
                FOR P:=0 TO SIZE DO
                    R[I,P+I]:=H[P]*G[I];
            FOR I:=0 TO KN DO H[I]:=0.0;
            SIZE:=SIZE+K;
            FOR P:=0 TO SIZE DO
                FOR I:=0 TO K DO
                    H[P]:=R[I,P]+H[P];
```

```
          END;
    FOR P:=0 TO KN DO G[P]:=H[P];
END;
BEGIN
{ Initialize array, N , with the values for n to be considered. }
    N[1]:=2;N[2]:=4;N[3]:=8;N[4]:=16;N[5]:=21;
    N[6]:=25;N[7]:=27;N[8]:=32;N[9]:=34;N[10]:=35;N[11]:=50;
    FOR TRIALS:=1 TO 11 DO
         BEGIN
              E:=0.0;
              FOR L:=1 TO K DO
              FOR I:=0 TO KN DO F[L][I]:=0.0;
         FOR I:=0 TO K DO
              F[K][I]:=BIN(K,I);
         FOR L:=K-1 DOWNTO 1 DO
              FOR I:=0 TO L DO
                   F[L][I]:=F[K][I];
         FOR L:=1 TO K DO
              POWER (F[L],N[TRIALS]);
         FOR I:=0 TO KN DO
              BEGIN
                   FOR L:=K-1 DOWNTO 1 DO
                   F[K][I]:=K*F[K][I];
                        F[K][I]:=F[K][I]-F[L][I];
              END;
         F[K][0]:=F[K][0]-1;
         FOR I:=0 TO K*N[TRIALS]-1 DO
              E:=E+F[K][K*N[TRIALS]-I]/BIN(K*N[TRIALS]-1,I);
```

```
            { To modify this program for the case of the half
            deck, j = nk/2, let the above loop run from I:=0 TO
            TRUNC(K*N[TRIALS]/2)-1. }
            E:=E/(K*N[TRIALS]);
            WRITELN ('THE EXPECTED # OF HITS FOR A DECK OF
            SIZE');
            WRITELN (N[TRIALS]:3,'*4':2,' IS ':4,E:5:2);
            WRITELN;
        END;
END.
```

References

L. Bloom, Language Development: Form and Function in Emerging Grammars, M.I.T. Press, Cambridge, MA, 1970.

R. Brown, A First Language: The Early Stages, Harvard University Press, Cambridge, MA, 1973.

B. T. Gardner and R. A. Gardner, Two-way communication with an infant chimpanzee, in A. Schrier and F. Stollnitz (eds.), Behavior of Nonhuman Primates (Vol. 4 (1971) 117–184), Academic Press, New York.

R. A. Gardner and B. T. Gardner, Comparative psychology and language acquisition, Ann. of the New York Academy of Sciences, 309 (1978) 37–76.

R. A. Gardner and B. T. Gardner, A vocabulary test for chimpanzees, J. of Comparative Psychology, 98 (1984) 381–404.

K. J. Hayes and C. Hayes, The intellectual development of a home-raised chimpanzee, Proceedings of the American Philosophical Society, 95 (1951) 105–109.

H. M. Jenkins, Measurement of stimulus control during discriminative operant conditioning, Psychological Bulletin, 64 (1965) 365–376.

W. N. Kellogg and L. A. Kellogg, The Ape and the Child, Hafner, New York, 1933.

R. C. Read, Card-guessing with information—a problem in probability, this MONTHLY, 69 (1962) 506–511.

J. Umiker-Sebeok and T. A. Sebeok, Questioning apes, in T. A. Sebeok and J. Umiker-Sebeok, (eds.) Speaking of apes (pp. 1–59), Plenum, New York, 1980.

---

172. MISCELLANEA

In a world in which the price of calculation continues to decrease rapidly, but the price of theorem proving continues to hold steady or increase, elementary economics indicates that we ought to spend a larger and larger fraction of our time on calculation.

Source: John W. Tukey, *The American Statistician*, 40 (1986) p. 74.

Editorial comment: !

ANSWER TO PHOTO ON PAGE 519

Karen Uhlenbeck.

A NOTE ON ALTERNATING SERIES IN SEVERAL DIMENSIONS

DAVID BORWEIN
Department of Mathematics, The University of Western Ontario, London, Ontario, Canada N6A 5B7

JONATHAN M. BORWEIN
Department of Mathematics, Statistics and Computing Science, Dalhousie University, Halifax, Nova Scotia, Canada B3H 3J5

1. Introduction. In the course of a study of chemical lattice sums [1] the authors considered sums such as

$$\sum{}'(-1)^{n+m+k}(n^2+m^2+k^2)^{-1/2}, \tag{1.1}$$

the summation being over all non-zero integer triples. Such "sums" occur naturally in the study of crystal potentials. For example, (1.1) is meant to measure the potential at the origin of an infinite cubic crystal with unit Coulomb charges at each integer lattice point. As such the sum is considered to represent an electrochemical constant (Madelung's constant) for sodium chloride. An excellent account of such lattice sums can be found in Glasser and Zucker's recent survey [3].

The series in (1.1) is not absolutely convergent and hence its sum is not order independent. Various possible orders suggest themselves. The chemical literature is somewhat vague on this point. As discussed in [1], it is really only appropriate to consider rectangular sums.

We shall consider alternating series of the form

$$\sum(-1)^{\overline{m}}f(\overline{m}), \tag{1.2}$$

where $\overline{m}=(m_1,m_2,\dots,m_N)$ ranges over $\mathbb{N}^N$, the N-fold product of non-negative integers, $(-1)^{\overline{m}}:=(-1)^{m_1+m_2+\cdots+m_N}$ and $f\colon \mathbb{N}^N\to\mathbb{R}$. For $s\colon\mathbb{N}^N\to\mathbb{R}$, $s(\overline{n})$ is said to *converge to a limit l as $\overline{n}$ increases in $\mathbb{N}^N$* if, given $\varepsilon>0$, there is an $\overline{m}$ in $\mathbb{N}^N$ such that

$$|s(\overline{n})-l|<\varepsilon \text{ whenever } \overline{n}\geqslant\overline{m},$$

the notation $\overline{n}\geqslant\overline{m}$ or $\overline{m}\leqslant\overline{n}$ meaning that $n_i\leqslant m_i$ for $i=1,2,\dots,N$. This is equivalent to convergence in the sense of Pringsheim, which requires that $s(\overline{n})\to l$ as $\min\limits_{1\leqslant i\leqslant N} n_i\to\infty$. For discussion of various concepts of convergence, see [2], [4], [5], [6], [7] and the references therein. We shall show that the sum of (1.2) exists under appropriate conditions if it is defined as the limit

David Borwein was born in Lithuania and raised in South Africa. He received a B.Sc. in electrical engineering from the University of Witwatersrand in 1945 and a B.Sc. in mathematics from that university in 1948. In 1950 he obtained a Ph.D. in mathematics from University College, London, England (under the supervision of Stephen Bosanquet) and a D.Sc. from the same university in 1960. In 1959 he was elected a Fellow of the Royal Society of Edinburgh. He taught in St. Andrews, Scotland, for thirteen years before moving to The University of Western Ontario, London, Canada, where he has served as Professor since 1963 and Head of the Department of Mathematics since 1967. He has written extensively on summability theory and other areas of analysis. He has served as Advanced Problems Editor for the MONTHLY and is President of the Canadian Mathematical Society. His hobbies include bridge and jogging. In addition to supervising eleven doctoral students, he has also reared two mathematicians: his sons Peter and Jonathan.

Jonathan Borwein was reared in Scotland and Canada. In 1971 he completed a B.Sc. in mathematics at The University of Western Ontario and won a Rhodes Scholarship to Oxford University where he obtained a D.Phil. under the supervision of M.A.H. Demptster in 1974. Since then he has taught at Dalhousie University, Halifax, Canada, where he is a Professor of Mathematics. He was also on the faculty at Carnegie-Mellon from 1980 to 1982. His principal research interests are optimization theory, analysis, and functional analysis. With his brother he has just completed a book on Analytic Computation Theory. He is an ex-competitive bridge player, a distance swimmer, and an election organizer.

This research was partially supported by the Natural Sciences and Engineering Research Council of Canada, and completed while the second author was visiting the Control and Management Systems Division of Cambridge University.

as $\bar{n}$ increases in $\mathbb{N}^N$ of

$$\sum_{\bar{0}\leqslant\bar{m}\leqslant\bar{n}} (-1)^{\bar{m}} f(\bar{m}). \tag{1.3}$$

Our main result, Theorem 3.1, is a complete generalization to multiple series of the classical result due to Leibniz, according to which if $\{a_n\}$ is a monotonic sequence of non-negative real numbers converging to zero, then the series $\sum_{n=0}^{\infty}(-1)^n a_n$ converges to s and its sequence of partial sums $\{s_n\}$ satisfies

$$0 \leqslant s_{2m+1} \leqslant s_{2m+3} \leqslant s \leqslant s_{2n+2} \leqslant s_{2n} \text{ for all } m, n \text{ in } \mathbb{N}.$$

Theorem 3.1 exhibits order relations between partial sums of the form (1.3) which hold when the function f is fully monotone (as defined in § 2 below); and Lemma 3.1, which plays a key role in the proof of Theorem 3.1, gives precise bounds on the size of these partial sums and also on the difference between them and their limit when it exists (see (3.6) below). To our knowledge neither such order nor such bound results have been considered previously in more than one dimension. There are, however, results available guaranteeing the convergence of non-absolutely convergent multiple dimensional series. Hardy [**4**] derives a "bounded convergence" test based on Abel partial summation which can be used to establish convergence criteria for series like (1.2) (see also Móricz [**6**]). Bromwich [**2**, p. 97] discusses the 2-dimensional version of Hardy's result. Meyer [**5**] gives necessary and sufficient conditions for the "diagonal summability" of 2-dimensional monotonic alternating series.

When f is the restriction of an N-times continuously differentiable function, a most satisfactory test involving partial derivatives is given by our Theorem 4.1 for the convergence of the alternating series (1.2). This theorem follows immediately from our Lemma 2.1 which has independent interest as a mean value estimate for general alternating sums.

We prove all our results for N-dimensions. The interested reader will be able to provide much simpler arguments for the 2-dimensional case as indicated pictorially in § 3 below. The 3-dimensional case, which is of primary interest, is not substantially simpler than the general one. Our vector notation enables us to express N-dimensional results concisely. An explicit treatment of the series (1.1) and its 2-dimensional analogue is given in [**1**].

2. Preliminaries. Let $\mathbb{N}^N, \mathbb{Z}^N, \mathbb{P}^N, \mathbb{R}^N$ denote respectively the N-fold product of the non-negative integers, the integers, the non-negative real numbers, and the real numbers. We denote by $\bar{1}$ the vector in $\mathbb{R}^N$ with every component 1, and by $\bar{e}_j$ the vector with jth component 1 and every other component 0. For

$$\bar{a} = (a_1, a_2, \dots, a_N), \bar{b} = (b_1, b_2, \dots, b_N) \text{ in } \mathbb{R}^N,$$

we define

$$|\bar{a}| := \max_{1\leqslant i\leqslant n} |a_i|, \quad \bar{a}\cdot\bar{b} := a_1 b_1 + a_2 b_2 + \cdots + a_n b_n, \quad \bar{a}\bar{b} := (a_1 b_1, a_2 b_2, \dots, a_N b_N);$$

and for $\bar{s} = (s_1, s_2, \dots, s_N)$ in $\mathbb{Z}^N$, we define

$$(-1)^{\bar{s}} := (-1)^{\bar{1}\cdot\bar{s}} = (-1)^{s_1+s_2+\cdots+s_N}.$$

A function $f\colon \mathbb{N}^N \to \mathbb{R}$ is said to be $(N-)$ *monotone* (*decreasing*) if

$$\sum_{\substack{|\bar{s}|\leqslant 1\\ \bar{s}\in\mathbb{N}^N}} (-1)^{\bar{s}} f(\bar{m}+\bar{s}) \geqslant 0 \tag{2.1}$$

for all $\bar{m}$ in $\mathbb{N}^N$. We shall say that f is *fully monotone* if f and all its coordinate restrictions are monotone. Then 1-monotonicity means that $f(m) \geqslant f(m+1)$ for m in $\mathbb{N}$; 2-monotonicity requires that

$$f(m,n)+f(m+1,n+1)\geqslant f(m,n+1)+f(m+1,n)\text{ for } m,n \text{ in } \mathbb{N};$$

and in general N-monotonicity of f requires the alternating sum (1.2) over any unit N-cube to have the same sign as $(-1)^{\bar{m}}$ at the corner nearest the origin. Note that any linear functional Φ on $\mathbb{R}^N$ is N-monotone for $N>1$, but Φ is fully monotone only if $\Phi\leqslant 0$ on $\mathbb{P}^N$.

We prove a lemma useful for obtaining criteria for monotonicity in which subscripts denote partial derivatives taken in order of the subscripts and

$$\mathbb{P}^N_*:=\{\bar{x}\in\mathbb{P}^N \mid x_i>0 \quad \text{for } i=1,2,\dots,N\}.$$

LEMMA 2.1. (a) *Let the function $f\colon \mathbb{P}^N_*\to\mathbb{R}$ have partial derivatives $f_{12\dots N}$ throughout $\mathbb{P}^N_*$. Then, for $\bar{x}\in\mathbb{P}^N_*$, $\bar{a}\in\mathbb{P}^N$,*

$$\sum_{\substack{|\bar{s}|\leqslant 1\\ \bar{s}\in\mathbb{N}^N}}(-1)^{\bar{s}}f(\bar{x}+\bar{a}\bar{s})=(-1)^N a_1a_2\cdots a_N f_{12\dots N}(\bar{c}) \tag{2.2}$$

for some $\bar{c}$ between $\bar{x}$ and $\bar{x}+\bar{a}$.

(b) *Let the function $\psi\colon\mathbb{P}_*\to\mathbb{R}$ be N-times differentiable. Then, for $x>0$, $\bar{a}\in\mathbb{P}^N$,*

$$\sum_{\substack{|\bar{s}|\leqslant 1\\ \bar{s}\in\mathbb{N}^N}}(-1)^{\bar{s}}\psi(x+\bar{a}\cdot\bar{s})=(-1)^N a_1a_2\cdots a_N\psi^{(N)}(c) \tag{2.3}$$

for some c between x and $x+a_1+a_2+\cdots+a_N$.

Proof. (a) The result is true for $N=1$. Suppose inductively that it is true with $N-1$ in place of N. The left-hand side of (2.2) is evidently equal to

$$\sum_{\substack{|\bar{s}|\leqslant 1\\ \bar{s}\in\mathbb{N}^{N-1}}}(-1)^{\bar{s}}g(x_1+a_1s_1,x_2+a_2s_2,\dots,x_{N-1}+a_{N-1}s_{N-1}),$$

where $g\colon\mathbb{P}^{N-1}_*\to\mathbb{R}$ is defined by

$$g(\bar{t}):=f(\bar{t},x_N)-f(\bar{t},x_N+a_N).$$

Hence, by the inductive hypothesis and the mean value theorem, the left-hand side of (2.2) equals

$$\begin{aligned}(-1)^{N-1}&a_1a_2\cdots a_{N-1}g_{12\dots N-1}(c_1,c_2,\dots,c_{N-1})\\ &=(-1)^N a_1a_2\cdots a_N f_{12\dots N}(c_1,c_2,\dots,c_N),\end{aligned}$$

where $x_i\leqslant c_i\leqslant x_i+a_i$ for $i=1,2,\dots,N$.

(b) Define $f\colon\mathbb{P}^N_*\to\mathbb{R}$ by $f(\bar{t}):=\psi(t_1+t_2+\cdots+t_N)$. Then, with $\bar{x}:=(x,x,\dots,x)/N$, we have

$$f(\bar{x}+\bar{a}\bar{s})=\psi(x+\bar{a}\cdot\bar{s});$$

and (2.3) follows from (2.2). □

The following lemma yields a stock of fully monotone functions.

LEMMA 2.2. *Let the function $\psi\colon\mathbb{P}\to\mathbb{R}$ be continuous and satisfy*

$$(-1)^n\psi^{(n)}(x)\geqslant 0 \quad \textit{for all } x>0 \quad \textit{and} \quad n=1,2,\dots,N. \tag{2.4}$$

Then

(a) *for $x\geqslant 0$, $\bar{a}\in\mathbb{P}^N$,*

$$\sum_{\substack{|\bar{s}|\leqslant 1\\ \bar{s}\in\mathbb{N}^N}}(-1)^{\bar{s}}\psi(x+\bar{a}\cdot\bar{s})\geqslant 0; \tag{2.5}$$

(b) *given non-decreasing functions* $g_i\colon \mathbb{N} \to \mathbb{N}$, $i = 1, 2, \ldots, N$, *the function* $f\colon \mathbb{N}^N \to \mathbb{R}$ *defined by*

$$f(\overline{m}) := \psi\left(\sum_{i=1}^{N} g_i(m_i)\right) \tag{2.6}$$

is fully monotone on $\mathbb{N}^N$.

Proof. (a) It follows from (2.3) that (2.5) holds for $x > 0$. By continuity it also holds for $x = 0$.

(b) We have

$$\sum_{|\bar{s}|\leqslant 1} (-1)^{\bar{s}} f(\overline{m} + \bar{s}) = \sum_{|\bar{s}|\leqslant 1} (-1)^{\bar{s}} \psi\left(\sum_{i=1}^{N} g_i(m_i) + \sum_{i=1}^{N} s_i(g_i(m_i + 1) - g_i(m_i))\right) \geqslant 0$$

by part (a), since $g_i(m_i + 1) - g_i(m_i) \geqslant 0$ and $\Sigma_{i=1}^{N} g_i(m_i) \geqslant 0$. Thus f in N-monotone. Since the argument applies with some of the g_i constant, it follows that f is in fact fully monotone. □

EXAMPLE 2.1. (a) Let

$$\|\overline{m}\|_p := \left(\sum_{i=1}^{N} |m_i|^p\right)^{1/p} \quad (p > 0).$$

Then

$$f(\overline{m}) := \|\bar{1} + \overline{m}\|_p^{-q} \quad (q > 0)$$

is fully monotone, by Lemma 2.2(b) with

$$\psi(x) := (N + x)^{-q/p}$$

and each $g_i(x) := (1 + x)^p - 1$. Note that with $\|\overline{m}\|_\infty := |\overline{m}|$ the function f is still fully monotone for $p = \infty$.

(b) Similarly the functions

$$f(\overline{m}) := -\log\left(\sum_{i=1}^{N} e^{a_i m_i}\right) \quad (a_i > 0)$$

and

$$f(\overline{m}) := \prod_{i=1}^{N} (1 + m_i)^{-1/N} = \exp\left(-\frac{1}{N}\sum_{i=1}^{N} \log(1 + m_i)\right)$$

are fully monotone. □

3. Alternating sums over rectangles. Before proceeding to the main results it is convenient to prove the following lemma.

LEMMA 3.1. *Let* $f\colon \mathbb{N}^N \to \mathbb{P}$ *be fully monotone. Then*

$$0 \leqslant (-1)^{\bar{a}} \sum_{\bar{a} \leqslant \overline{m} \leqslant \bar{n}} (-1)^{\overline{m}} f(\overline{m}) \leqslant f(\bar{a}) \tag{3.1}$$

whenever $\bar{a}, \bar{n} \in \mathbb{N}^N$ *and* $\bar{a} \leqslant \bar{n}$.

Proof. It suffices to prove (3.1) for the case $\bar{a} = \bar{0}$, since the general case follows from this case with $f(\overline{m})$ replaced by $f(\overline{m} + \bar{a})$. We establish the first inequality in (3.1) by induction. Clearly it holds for $N = 1$. Suppose it holds with $N - 1$ in place of N. Observe that, for $\bar{n} = (n_1, n_2, \ldots, n_N) \in \mathbb{N}^N$,

$$\sum_{\bar{0} \leqslant \bar{m} \leqslant \bar{n}} (-1)^{\bar{m}} f(\bar{m}) = S_1 + S_2$$

with

$$S_1 := \sum_{\bar{0} \leqslant 2\bar{k} \leqslant \bar{n}} (-1)^{2\bar{k}} \sum_{\bar{0} \leqslant \bar{s} \leqslant \bar{1}} (-1)^{\bar{s}} f(2\bar{k} + \bar{s}),$$

$$S_2 := \sum_{i \in E} \sum_{\substack{m_i = n_i \\ \bar{0} \leqslant \bar{m} \leqslant \bar{a}_i}} (-1)^{\bar{m}} f(\bar{m}),$$

where $E := \{i | n_i \text{ even}\}$, and $\bar{a}_i := (\alpha_1, \alpha_2, \dots, \alpha_N)$ with

$$\alpha_j := \begin{cases} n_j - 1 & \text{if } j < i,\ j \in E \text{ and } n_j \neq 0, \\ n_j & \text{otherwise.} \end{cases}$$

Now $S_1 \geqslant 0$ because of the N-monotonicity of f and $S_2 \geqslant 0$ by the inductive hypothesis. The validity of the first inequality in (3.1) for every $N \geqslant 1$ follows.

To establish the second inequality in (3.1), we observe that

$$\sum_{\bar{0} \leqslant \bar{m} \leqslant \bar{n}} (-1)^{\bar{m}} f(\bar{m}) - f(\bar{0}) = - \sum_{\substack{\bar{0} \leqslant \bar{s} \leqslant 1 \\ \bar{s} \neq \bar{0}}} (-1)^{\bar{s}} \sum_{\bar{s} \leqslant \bar{m} \leqslant \bar{n}} (-1)^{\bar{m}} f(\bar{m}) \leqslant 0$$

by the first inequality in (3.1). □

In what follows we use the notation:

$$t_{\bar{n}} := t_{\bar{n}}(f) := \sum_{\bar{0} \leqslant \bar{m} \leqslant \bar{n}} (-1)^{\bar{m}} f(\bar{m}), \tag{3.2}$$

where $\bar{m}, \bar{n} \in \mathbb{N}^N$.

THEOREM 3.1 (Alternating series test). *If $f\colon \mathbb{N}^N \to \mathbb{P}$ is fully monotone, $\bar{n}, \bar{r} \in \mathbb{N}^N$ and $\bar{n} \leqslant \bar{r}$, then*

(i) $t_{2\bar{n}} \geqslant t_{2\bar{r}} \geqslant 0$,
(ii) $t_{2\bar{r}+\bar{1}} \geqslant t_{2\bar{n}+\bar{1}} \geqslant 0$,
(iii) $t_{2\bar{n}} \geqslant t_{2\bar{n}+\bar{1}}$.

If in addition $\lim_{n \to \infty} f(n\bar{e}_i) = 0$ *for* $i = 1, 2, \dots, N$, *then* $t_{\bar{n}}$ *converges to a limit* t *as* $\bar{n}$ *increases,* $t_{2\bar{n}} \geqslant t \geqslant t_{2\bar{n}+\bar{1}}$ *for each* $\bar{n} \in \mathbb{N}^N$, *and consequently*

$$t = \sum_{m_1=0}^{\infty} \sum_{m_2=0}^{\infty} \cdots \sum_{m_N=0}^{\infty} (-1)^{\bar{m}} f(\bar{m}). \tag{3.3}$$

Proof. To establish (i) and (ii) it suffices to consider $\bar{r} := \bar{n} + \bar{e}_j, 1 \leqslant j \leqslant N$. We then have

$$t_{2\bar{r}} - t_{2\bar{n}} = \sum_{\bar{a} \leqslant \bar{m} \leqslant 2\bar{r}} (-1)^{\bar{m}} f(\bar{m}) \quad \text{with } \bar{a} := (2n_j + 1)\bar{e}_j,$$

and

$$t_{2\bar{r}+\bar{1}} - t_{2\bar{n}+\bar{1}} = \sum_{\bar{b} \leqslant \bar{m} \leqslant 2\bar{r}+\bar{1}} (-1)^{\bar{m}} f(\bar{m}) \quad \text{with } \bar{b} := (2n_j + 2)\bar{e}_j.$$

Since $(-1)^{\bar{a}} = -1$ and $(-1)^{\bar{b}} = 1$, (i) and (ii) follow by Lemma 3.1. Next, by an "inclusion-exclusion" counting argument, we have

$$t_{2\bar{n}+\bar{1}} - t_{2\bar{n}} = - \sum_{i=1}^{2^N-1} (-1)^{\bar{a}_i} \sum_{\bar{a}_i \leqslant \bar{m} \leqslant 2\bar{n}+\bar{1}} (-1)^{\bar{m}} f(\bar{m}),$$

each $\bar{a}_i$ being of the form $(\alpha_1, \alpha_2, \ldots, \alpha_N) \neq \bar{0}$ with every α_j either 0 or $2n_j + 1$. Lemma 3.1 now yields (iii).

It follows from (i) that $t_{2\bar{n}}$ converges (to $\inf_{\bar{n} \in \mathbb{N}^N} t_{2\bar{n}}$) as $\bar{n}$ increases in $\mathbb{N}^N$. Hence to establish the convergence of $t_{\bar{n}}$ as $\bar{n}$ increases in $\mathbb{N}^N$, it suffices to show that, for fixed $\bar{k}$ in $\mathbb{N}^N$, $t_{\bar{n}+\bar{k}} - t_{\bar{n}} \to 0$ as $\bar{n}$ increases, and for this it is enough to consider $\bar{k} := \bar{e}_j$. We then have

$$t_{\bar{n}+\bar{k}} - t_{\bar{n}} = \sum_{\bar{c} \leqslant \bar{m} \leqslant \bar{n}+\bar{k}} (-1)^{\bar{m}} f(\bar{m}) \quad \text{with } \bar{c} := (n_j + 1)\bar{e}_j,$$

and hence, by Lemma 3.1,

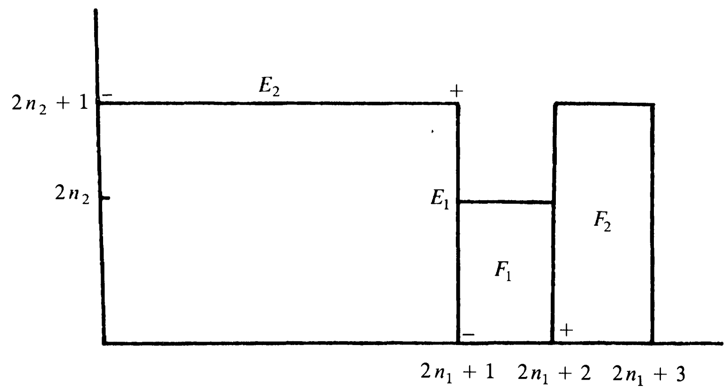
$$|t_{\bar{n}+\bar{k}} - t_{\bar{n}}| \leqslant f(\bar{c}) \tag{3.4}$$

which tends to 0 as $\bar{n}$ increases.

Since f is fully monotone, an inductive argument shows that the sums implicit in (3.3) exist and that the identity holds. □

The convergence conclusion also follows from Hardy's "bounded series" test which also requires full monotonicity without so naming it. Hardy's result, however, does not yield the alternation information expressed by (i), (ii) and (iii).

To indicate the underlying geometric simplicity of our proof we show pictorially why (i), (ii) and (iii) hold in the two-dimensional case when $\bar{n} = (n_1, n_2)$ and $\bar{r} = (n_1 + 1, n_2)$.



$$t_{2\bar{r}} - t_{2\bar{n}} = F_1 \leqslant 0;$$
$$t_{2\bar{r}+\bar{1}} - t_{2\bar{n}+\bar{1}} = F_2 \geqslant 0;$$
$$t_{2\bar{n}+\bar{1}} - t_{2\bar{n}} = E_1 + E_2 - f(2\bar{n} + \bar{1})$$
$$E_1 \leqslant 0, E_2 \leqslant 0.$$

Here F_i represents the sum over the enclosed rectangle and E_i the sum over the adjacent edge.

The next theorem is a special case of Theorem 3.1. We use the notation:

$$t_n := t_n(f) := \sum_{|\bar{m}| \leqslant n} (-1)^{\bar{m}} f(\bar{m}), \tag{3.5}$$

where $\bar{m} \in \mathbb{N}^N$ and $n \in \mathbb{N}$. Observe that $t_n = t_{n\bar{1}}$.

THEOREM 3.2. *If $f\colon \mathbb{N}^N \to \mathbb{P}$ is fully monotone, then, for $m, n \in \mathbb{N}$,*

$$0 \leqslant t_{2m+1} \leqslant t_{2m+3} \leqslant t_{2n+2} \leqslant t_{2n}.$$

If in addition $\lim_{n\to\infty} f(n\bar{e}_i) = 0$ for $i = 1, 2, \ldots, N$, then t_n converges to a limit t as $n \to \infty$, and $t_{2n} \geqslant t \geqslant t_{2n+1}$ for each $n \in \mathbb{N}$.

COROLLARY 3.3. *For* $0 < p \leqslant \infty, q > 0, \bar{n} \in \mathbb{N}^N$,

$$S_{\bar{n}}^N := \sum_{\bar{1} \leqslant \bar{m} \leqslant \bar{n}} (-1)^{\overline{m}} \|\bar{m}\|_p^{-q}$$

converges as $\bar{n}$ *increases.*

Proof. We have

$$S_{\bar{n}+\bar{1}}^N = (-1)^N \sum_{\bar{0} \leqslant \bar{m} \leqslant \bar{n}} (-1)^{\overline{m}} \|\bar{1} + \bar{m}\|_p^{-q}$$

and Theorem 3.1 applied to Example 2.1 (a) yields the required convergence. □

COROLLARY 3.4. *For* $0 < p \leqslant \infty, q > 0, \bar{n} \in \mathbb{N}^N$,

$$A_{\bar{n}}^N := \sum_{\substack{-\bar{n} \leqslant \bar{m} \leqslant \bar{n} \\ \bar{m} \in \mathbb{Z}^N \setminus \{\bar{0}\}}} (-1)^{\overline{m}} \|\bar{m}\|_p^{-q}$$

converges as $\bar{n}$ *increases.*

Proof. Observe that

$$A_{\bar{n}}^N = 2^N S_{\bar{n}}^N + R,$$

where R is the sum of a number of finite series each of the same general form as $A_{\bar{n}}^N$ but of lower dimension. The desired result follows by induction. □

Similarly we have:

COROLLARY 3.5. *For* $0 < p \leqslant \infty, q > 0, \bar{a} \in \mathbb{R}^N, \bar{n} \in \mathbb{N}^N$,

$$\sum_{\substack{-\bar{n} \leqslant \bar{m} \leqslant \bar{n} \\ \bar{m} \in \mathbb{Z}^N \setminus \{\bar{0}\}}} (-1)^{\overline{m}} \|\bar{a}\bar{m}\|_p^{-q}$$

converges as $\bar{n}$ *increases.*

In particular Madelung's constant exists for any rectilinear lattice in $\mathbb{R}^N$ if defined as

$$M^N(\bar{a}) := \lim_{n \to \infty} \sum_{\substack{|\bar{m}| \leqslant n \\ \bar{m} \in \mathbb{Z}^N \setminus \{\bar{0}\}}} (-1)^{\overline{m}} \|\bar{a}\bar{m}\|_2^{-1}.$$

By virtue of the underlying alternation it is easy to obtain a good error bound for $t - t_{\bar{n}}$, when $t, t_{\bar{n}}$ are as in Theorem 3.1. If $f\colon \mathbb{N}^N \to \mathbb{R}$ is fully monotone and $t_{\bar{n}}$ converges to t as $\bar{n}$ increases, then, letting $\bar{n}_0 := \bar{n}$ and $\bar{n}_j := \bar{n}_{j-1} + \bar{e}_j$ for $j = 1, 2, \ldots, N$, we have

$$|t - t_{\bar{n}}| \leqslant |t_{\bar{n}+\bar{1}} - t_{\bar{n}}| \leqslant \sum_{j=1}^{N} |t_{\bar{n}_j} - t_{\bar{n}_{j-1}}| \leqslant \sum_{j=1}^{N} f\big((n_j + 1)\bar{e}_j\big), \tag{3.6}$$

on repeated application of (3.4). For the series in (1.1) the difference between

$$\sum_{\substack{|\bar{m}| \leqslant n \\ \bar{m} \in \mathbb{Z}^N \setminus \{\bar{0}\}}} (-1)^{\overline{m}} \|\bar{m}\|_2^{-1}$$

and the limit is at most $N/(n+1)$. Thus, to compute 15 digits of Madelung's constant for NaCl directly would appear to take around 10^{45} calculations! No wonder indirect transform techniques are used in practice [3]. Actually only a few digits seem to be used in applications. The NaCl crystal would have to be galaxy sized for a 15 digit approximation to have physical significance. This indicates the limited utility of using an infinite model for a finite crystal.

By virtue of Example 2.1 (b) and Theorem 3.1,

$$\sum_{\bar{1}\leqslant\bar{m}\leqslant\bar{n}}(-1)^{\bar{m}}\left(\prod_{i=1}^{N}m_i\right)^{-1/N}$$

converges as $\bar{n}$ increases in $\mathbb{N}^N$. This is the limiting case as $p \to 0$ of Corollary 3.3 with $q = 1$. In fact the sum is just

$$\left(\sum_{m=1}^{\infty}(-1)^m m^{-1/N}\right)^N.$$

More generally, in view of Lemma 2.2, Theorem 3.1 applies to $f(\bar{m}):=\prod_{i=1}^{N}g_i(m_i)$ whenever each g_i: $\mathbb{N} \to \mathbb{N}$ decreases to zero. In this case, of course,

$$\sum_{\bar{0}\leqslant\bar{m}\leqslant\bar{n}}(-1)^{\bar{m}}\prod_{i=1}^{N}g_i(m_i)$$

converges to

$$\prod_{i=1}^{N}\sum_{m=0}^{\infty}(-1)^m g_i(m)$$

as $\bar{n}$ increases in $\mathbb{N}^N$.

4. A characterization involving partial derivatives. We define a function f: $\mathbb{P}^N \to R$ to be $(N\text{-})$*monotone* if

$$\sum_{\substack{|\bar{s}|\leqslant 1\\ \bar{s}\in\mathbb{N}^N}}(-1)^{\bar{s}}f(\bar{x}+\bar{a}\bar{s}) \geqslant 0 \tag{4.1}$$

whenever $\bar{x}, \bar{a} \in \mathbb{P}^N$. We shall say that f is *fully monotone* on $\mathbb{P}^N$ if f and all its coordinate restrictions are monotone. Then Lemma 2.2 (b) has an obvious analogue. More importantly, we have the following characterization of full monotonicity on $\mathbb{P}^N$ in which subscripts denote partial derivatives.

THEOREM 4.1. *Let f: $\mathbb{P}^N \to \mathbb{R}$ have continuous partial derivatives of order N. Then*
(a) f is monotone on $\mathbb{P}^N$ if and only if $(-1)^N f_{12\ldots N} \geqslant 0$, and
(b) f is fully monotone on $\mathbb{P}^N$ if and only if, for $1 \leqslant k \leqslant N$,

$$(-1)^k f_{i_1 i_2\ldots i_k} \geqslant 0 \tag{4.2}$$

whenever $i_1, i_2, \ldots, i_k$ are distinct integers in $\{1,2,\ldots,N\}$.

Proof. (a) By Lemma 2.1 (a), for $\bar{x} \in \mathbb{P}_*^N$, $\bar{a} \in \mathbb{P}^N$,

$$\sum_{|\bar{s}|\leqslant 1}(-1)^{\bar{s}}f(\bar{x}+\bar{a}\bar{s}) = (-1)^N a_1 a_2 \cdots a_N f_{12\ldots N}(\bar{c}), \tag{4.3}$$

where $\bar{x} \leqslant \bar{c} \leqslant \bar{x} + \bar{a}$. In view of the continuity of f and $f_{12\ldots N}$ on $\mathbb{P}^N$, conclusion (a) follows from (4.3).

(b) This follows from (a) by consideration of coordinate restrictions. □

Note that (4.3) shows that if $(-1)^N f_{12\ldots N} > 0$ on $\mathbb{P}^N$, then f is strictly N-monotone. Further, since full monotonicity on $\mathbb{P}^N$ implies full monotonicity on $\mathbb{N}^N$, (4.2) yields a simple test for full monotonicity on $\mathbb{N}^N$. Let $\mathbb{R}_{\bar{c}}^N := \{\bar{x} \in \mathbb{R}^N | \bar{x} \geqslant \bar{c}\}$ and say that f is N-monotone or fully monotone on $\mathbb{R}_{\bar{c}}^N$ if $f(\bar{x}-\bar{c})$ is similarly monotone on $\mathbb{R}_{\bar{0}}^N = \mathbb{P}^N$.

EXAMPLE 4.2. (a) Let $f(x,y,z) := (x^2+y^2+z^2)^{-q}$ for $q > 0$ and $(x,y,z) \in \mathbb{R}_{\bar{1}}^3$. Then

$$f_1 = -2xq(x^2+y^2+z^2)^{-q-1} < 0, \quad f_{12} = 4xyq(q+1)(x^2+y^2+z^2)^{-q-2} > 0,$$

and

$$f_{123} = -8xyzq(q+1)(q+2)(x^2+y^2+z^2)^{-q-3} < 0$$

on $\mathbb{R}^3_1$. By symmetry this suffices to show that f is (strictly) fully monotone on $\mathbb{R}^3_1$. The N-dimensional case can be treated similarly, as can the function in Example 2.1 (a).

(b) Consider the two-dimensional lattice sum

$$\sum_{m,n=1}^{\infty} (-1)^{m+n}\left((a_1m+b_1n)^2+(a_2m+b_2n)^2\right)^{-q}, \tag{4.4}$$

where $q > 0$ and a_1, a_2, b_1, b_2 are real numbers. This corresponds to summing over the cone of vectors of the form $m\bar{a} + n\bar{b}$ for $m \geqslant 1$, $n \geqslant 1$, where $\bar{a} := (a_1, a_2)$, $\bar{b} := (b_1, b_2)$. Let

$$A := a_1^2 + a_2^2, \quad B := b_1^2 + b_2^2, C := \bar{a}\cdot\bar{b} = a_1b_1 + a_2b_2,$$

and

$$D := D(x,y) = (a_1x+b_1y)^2 + (a_2x+b_2y)^2.$$

Then $f(x,y) := D^{-q}(x,y)$ satisfies

$$f_1 = -2q(Ax+Cy)D^{-q-1}, \quad f_2 = -2q(Cx+By)D^{q-1},$$

$$f_{12} = 4q(q+1)(Ax+Cy)(Cx+By)D^{-q-2} - 2qCD^{-q-1}$$

$$= 2qD^{-q-2}\left((2q+1)(Ax^2+By^2)C + 2xy\left(qC^2+(q+1)AB\right)\right).$$

We see that $f_{12} \geqslant 0$ while $f_1 \leqslant 0$ and $f_2 \leqslant 0$ on $\mathbb{R}^2_1$ if and only if $C \geqslant 0$. Thus f is fully monotone on $\mathbb{R}^2_1$ if and only if the angle between $\bar{a}$ and $\bar{b}$ is acute, and in this case the series in (4.4) converges in the sense of Theorem 3.1 or Theorem 3.2. □

References

1. D. Borwein, J. M. Borwein, and K. F. Taylor, Convergence of lattice sums and Madelung's constant, J. Math. Phys., 26 (1985) 2999–3009.
2. T. J. Bromwich, An Introduction to the Theory of Infinite Series, 2nd ed., Macmillan, 1926.
3. M. L. Glasser and I. J. Zucker, Lattice sums, in Perspectives in Theoretical Chemistry: Advances and Perspectives, 5(1980) 67–139.
4. G. H. Hardy, On the convergence of certain multiple series, Proc. London Math. Soc., 2(1904) 24–28 and Proc. Cambridge Philos. Soc., 19(1917) 86–95. (See also Collected Papers, vol. 6, pages 5–10, and 629–639, Oxford, 1974.)
5. B. Meyer, On the convergence of alternating double series, this MONTHLY, 60(1953) 402–404.
6. F. Móricz, Some remarks on the notion of regular convergence of multiple series, Acta Math. Hungar., 41(1983) 161–168.
7. A. Wilansky, On the convergence of double series, Bull. Amer. Math. Soc., 53(1947) 793–799.

---

173. **MISCELLANEA**

Too long for haiku!
Truly marvellous the proof
Of Fermat's theorem.

Robert J. MacG. Dawson
Corpus Christi College
Cambridge, United Kingdom
CB2 1RH

INFINITESIMALS, MICROSIMPLEXES AND ELEMENTARY HOMOLOGY THEORY

RADE T. ŽIVALJEVIĆ
Mathematics Institute, Knez Mihailova 35/1, 11000 Beograd, Yugoslavia

The purpose of this article is to draw the attention of the general mathematical audience to a not very well known application of Abraham Robinson's infinitesimal analysis or nonstandard analysis as it is more often called. The article should be easily accessible to anybody who has just a basic acquaintance with the method of infinitesimals, particularly to those advanced undergraduate students who have been taught by H. J. Keisler's book "Elementary Calculus" [3]. We plan, inspired by M. C. McCord's original paper "Nonstandard analysis and homology" [7], to motivate and give an exposition of some basic results of elementary homology theory including the proof of the Brouwer Fixed Point Theorem. The exposition is based on a naturally defined notion of a microsimplex, cf. [8], and a hyperfinite chain of microsimplexes which are derived from infinitesimal analysis.

The notation and terminology of nonstandard analysis used here is standard and agrees with that used in J. H. Keisler's books [3], [5]; what follows is a short reminder. It is well known by now that a closer look at any finite set S reveals its second nature, $^*S$. The set of real $\mathbb{R}$ is just the "visible" part of the set $^*\mathbb{R}$ which contains also infinitesimals, hyperfinite numbers, etc. For $x, y \in {}^*(\mathbb{R}^n)$, particularly for two *real numbers, $x \approx y$ means that the distance between x and y is infinitesimal. The monad of x in $\mathbb{R}^n$ is just the set $m(x) = \{y \in {}^*\mathbb{R}^n | y \approx x\}$ and the standard part map, or relation, is defined by $y = \mathrm{st}(x) \Leftrightarrow x \in m(y)$.

1. Brouwer's fixed-point theorem. Brouwer's celebrated theorem states that every continuous function $f: B^n \to B^n$, where B^n is the unit ball in n-dimensional Euclidean space $\mathbb{R}^n$, has a fixed point, i.e., $f(x) = x$ for some x in B^n. Recall the well-known fact, actually used as the definition of continuity in nonstandard calculus textbooks, that a function f defined on a subset A of $\mathbb{R}^n$ with values in $\mathbb{R}^m$ is continuous at a point a in A if and only if $\forall x\ (x \in {}^*A$ and $x \approx a \Rightarrow {}^*f(x) \approx {}^*f(a))$. To motivate the definition of a microsimplex and a hyperfinite chain of microsimplexes, let us review a nonstandard proof of the Brouwer fixed-point theorem (BFPT) in case $n = 1$ (cf. [3] p. 165 or [5] p. 81).

Proof of BFPT *in case* $n = 1$. Let $f: [0,1] \to [0,1]$ be a continuous function. Let H in $^*\mathbb{N} \setminus \mathbb{N}$ be a hyperfinite number and $(a_i)_{i=0}^{H}$ the sequence defined by $a_i = i/H \in {}^*[0,1]$. One can also say that the sequence $((a_i, a_{i+1}))_{i=0}^{H-1}$ spans or approximates the interval $[0,1]$. Now, since all objects are internal, which means that they are governed by the same laws which hold in the standard universe (Leibniz's Principle, cf. [5] p. 42), we conclude that there exists $i \leqslant H-1$ such that $^*f(a_i) \geqslant a_i$ and $^*f(a_{i+1}) \leqslant a_{i+1}$. Actually, one can take $i = \max\{j | a_j \leqslant {}^*f(a_j)$ and $j \leqslant H-1\}$. Let $x = \mathrm{st}(a_i) = \mathrm{st}(a_{i+1})$. From the nonstandard characterization of continuity it follows that

$$f(x) \approx {}^*f(a_i) \geqslant a_i \approx x \quad \text{and} \quad f(x) \approx {}^*f(a_{i+1}) \leqslant a_{i+1} \approx x \Rightarrow f(x) = x,$$

so x is a fixed point of f.

Unfortunately, the proof of the Brouwer theorem in higher dimensions is harder and a standard intermediate step is the following proposition.

R. T. Živaljević was born in 1954 in Sarajevo, Yugoslavia. He studied at the University of Belgrade, where he obtained both his Master's degree in combinatorial geometry and Ph.D. degree in nonstandard analysis under the supervision of Milosav Marjanović. His second (and last) Ph.D. degree (1985) was obtained in nonstandard analysis under the supervision of H. Jerome Keisler from the University of Wisconsin, Madison. Generally interested in applications of nonstandard analysis, he is a secret admirer of algebraic topology.

PROPOSITION 1. *If there exists a continuous function $f: B^n \to B^n$ without fixed points, then there exists a continuous function, called a retraction, $r: B^n \to S^{n-1}$ where S^{n-1} is the unit sphere, such that for each $x \in S^{n-1} r(x) = x$.*

Proof. Let $[f(x), x) := \{(1-\lambda)f(x) + \lambda x \mid \lambda \geqslant 0\}$ be the ray having $f(x)$ as the end-point which passes through x and let

$$\{r(x)\} = [f(x), x) \cap S^{n-1}.$$

The function r: $B^n \to S^{n-1}$ is obviously well defined, so to prove that it is continuous let us assume that $x \in B^n$ and $y \approx x$. Hence, $^*f(y) \approx f(x)$ which means that for some $\lambda_0 \in {}^*(\mathbb{R}_+)$

$$\mathrm{st}(^*r(y)) = \mathrm{st}((1-\lambda_0)^*f(y) + \lambda_0 y) = (1-\lambda)f(x) + \lambda x \quad \text{where} \quad \lambda = \mathrm{st}(\lambda_0).$$

Hence, $\mathrm{st}(^*r(y)) \in [f(x), x) \cap S^{n-1}$, i.e., $^*r(y) \approx r(x)$. On the other hand one can easily check that $r(x) = x$ for each $x \in S^{n-1}$.

To prove such an "obvious" fact that the retraction from the proposition above cannot exist, we need a notion of an infinitesimal simplex together with the combinatorics or elementary algebra of hyperfinite sequences of such simplexes.

DEFINITION 1. A p-dimensional microsimplex in a space $K \subset R^n$, or more generally in any topological space, is an ordered $(p+1)$-tuple $(a_0, a_1, \ldots, a_p)$ of elements in $^*K$ such that there exists $x \in K$ with the property $x = \mathrm{st}(a_i)$ for all $0 \leqslant i \leqslant p$. Let $M_p(K)$ be the set of all p-dimensional microsimplexes.

Obviously, 0-dimensional simplexes are just near standard points in the space and in general p-dimensional microsimplexes could be understood as elementary pieces of the space which can, if properly arranged, span or approximate in a certain sense any given p-dimensional part of the space. To achieve this approximation, these microsimplexes should be "glued" to each other to make a kind of tiling of the space. To formalize this we need the following definition.

DEFINITION 2 (M. C. McCord [7]). Let $(s_i)_{i=1}^H$ be an internal sequence of p-dimensional microsimplexes in a space $K \subset \mathbb{R}^n$, where $s_i = (a_0^i, a_1^i, \ldots, a_p^i)$, and $(n_i)_{i=1}^H$ an internal sequence of nonstandard integers. The pair $((s_i)_{i=1}^H, (n_i)_{i=1}^H)$ is called a hyperfinite chain of microsimplexes and will be formally denoted by $\sum_{i=1}^H n_i s_i$ which indicates that it should be thought of as a "formal sum". Therefore, two hyperfinite chains of microsimplexes, or shortly chains, are supposed to be equal if one can be obtained from the other by a convenient rearrangement of microsimplexes. To make this precise let $\langle s, u\rangle := {}^*\Sigma\{n_i | s_i = u\}$ where $s = \sum_{i=1}^H n_i s_i$ and $u \in M_p(K)$ and $^*\Sigma$ is just the internal sum of a hyperfinite set of hyperintegers. Then two chains $s = \sum_{i=1}^H n_i s_i$ and $t = \sum_{i=1}^D m_i t_i$ are said to be equal if $\langle s, u\rangle = \langle t, u\rangle$ for any $u \in M_p(K)$. This permits us also to define the sum of p-dimensional chains by the formula $\langle s+t, u\rangle = \langle s, u\rangle + \langle t, u\rangle$ and it is easily checked that $s+t$ really exists and does not depend on particular representatives.

PROPOSITION 2. *The (external) set $C_p(K)$ of all p-dimensional chains is an Abelian group with respect to addition of chains.*

Proof. The proof is easy and is left to the reader.

DEFINITION 3. Let $u = (u_0, u_1, \ldots, u_p)$ be a p-dimensional microsimplex. Then,

$$\partial_p u := \sum_{i=0}^{p} (-1)^i (u_0, \ldots, \hat{u}_i, \ldots, u_p)$$

is an element of $C_{p-1}(K)$ which is called the boundary of u. As usual $\hat{u}_i$ means that this term has been omitted. More generally, there is a homomorphism $\partial_p : C_p(K) \to C_{p-1}(K)$, called the boundary homomorphism, defined by

$$\partial_p\left(\sum_{i=1}^{H} n_i s_i\right) = \sum_{i=1}^{H} n_i \partial_p s_i .$$

A hyperfinite chain $s = \sum_{i=1}^{H} n_i s_i \in C_p(K)$ is called a cycle or a p-dimensional cycle if $\partial_p(s) = 0$, while we call it a boundary if $s = \partial_{p+1}(s')$ for some $s' \in C_{p+1}(K)$. The sets of all p-dimensional cycles and boundaries are denoted by $Z_p(K)$ and $B_p(K)$ respectively.

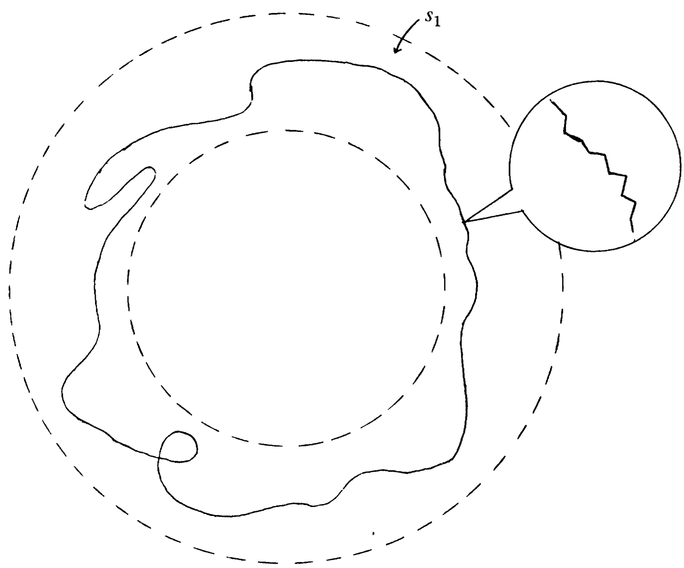


FIG. 1

It is easy to check that both $Z_p(K)$ and $B_p(K)$ are subgroups of $C_p(K)$. Informally, one can see p-cycles as those chains which are capable of approximating or surrounding $(p + 1)$-dimensional holes in the space K; see Fig. 1, where a 1-dimensional cycle s_1 detected the hole of the annulus. A typical boundary, say

$$\partial(a_0, a_1, a_2) = (a_1, a_2) - (a_0, a_2) + (a_0, a_1),$$

where $a = (a_0, a_1, a_2)$ is a microsimplex, is always a cycle. Indeed, one can easily check that $\partial_{p-1}\partial_p(a_0, a_1, \dots, a_p) = 0$, so by Definition 3 $\partial_{p-1}\partial_p(s) = 0$ for any $s \in C_p(K)$. However, if we are interested in "real" holes in the space K, those cycles which are boundaries should be ignored. Why? Well, if $c = \partial d$, then the portion of the space surrounded by c is obviously covered by d so it cannot be a "real" hole. To make this precise we need one more definition.

DEFINITION 4. Two p-dimensional cycles c_1 and c_2, or more generally two chains, are said to be homologous if $c_1 - c_2$ is a boundary (i.e., if they are cycles, they are surrounding the same hole!).

Obviously, a chain is homologous to 0 if and only if it is a boundary.

PROPOSITION 3. *Let $f: K \to L$ be a continuous map. Then f induces a homomorphism $f_\#^p : C_p(K) \to C_p(L)$ where*

$$f_\#^p\left(\sum_{i=1}^{H} n_i\left(a_0^i, \dots, a_p^i\right)\right) := \sum_{i=1}^{H} n_i\left({}^*fa_0^i, \dots, {}^*fa_0^i\right).$$

Furthermore, $\partial_p f_\#^p(s) = f_\#^{p-1}\partial_p(s)$ for $s \in C_p(K)$ which means that $f_\#^p$ sends cycles to cycles, boundaries to boundaries and homologous chains to homologous chains.

Proof. Since $f: K \to L$ is continuous, it sends monads of K into monads of L and microsimplexes of K to microsimplexes of L, which means that $f_{\#}^{p}$ is well-defined. The rest of the proof is an easy exercise.

Proof of BFPT *in case* $n \geqslant 2$. By Proposition 1 it would be enough to prove that there does not exist a retraction $r: B^n \to S^{n-1}$. In order to avoid awkward formulas and definitions, let us prove Brouwer's theorem assuming $n = 2$ first and then point to the obvious changes in the proof for the general case. The idea of the proof is as simple as possible. There is a hole in the circle S^1 and no holes in the ball B^2. In other words:

(a) there exists a 1-dimensional cycle $s \in C_1(S^1)$ which is not a boundary and

(b) any cycle $s' \in C_1(B^2)$ is a boundary!

Let us prove statements (a) and (b).

(a) Let $(a_i)_{i=0}^{H-1}$ be the sequence of consecutive vertices of a regular polygon, say

$$a_i = \left(\cos\frac{2\pi i}{H}, \sin\frac{2\pi i}{H}\right), H \in {}^*\mathbb{N} \setminus \mathbb{N}.$$

Then $s = \Sigma_{i=0}^{H-1}(a_i, a_{i+1})$ where $a_H := a_0$ is obviously a cycle which is not a boundary. To prove this let $\theta(a, b)$, where (a, b) is a 1-dimensional microsimplex, be the oriented angle between rays $[0, a)$ and $[0, b)$. Obviously θ can be extended to a homomorphism $\theta: C_1(S^1) \to {}^*\mathbb{R}$ by the formula

$$\theta\left(\sum_{i=1}^{D} n_i(a_i, b_i)\right) := {}^*\sum_{i=1}^{D} n_i\theta(a_i, b_i).$$

It is clear that $\theta(s) = 2\pi$ while $\theta(s') = 0$ for any boundary s'. Hence s is a cycle which is not a boundary.

(b) Let $s' = \Sigma_{i=1}^{D} n_i(a_i, b_i) \in C_1(B^2)$ be a cycle. Let $\varepsilon = 1/H$ for $H \in {}^*\mathbb{N} \setminus \mathbb{N}$. We claim that s' and $s'' = \Sigma_{i=1}^{D} n_i(c_i, d_i)$, where $c_i = (1-\varepsilon)a_i$, $d_i = (1-\varepsilon)b_i$ are homologous. Indeed, if $d_i = (a_i, b_i, c_i) + (b_i, d_i, c_i)$, then

$$\partial d_i = (a_i, b_i) + (d_i, c_i) + (b_i, d_i) - (a_i, c_i).$$

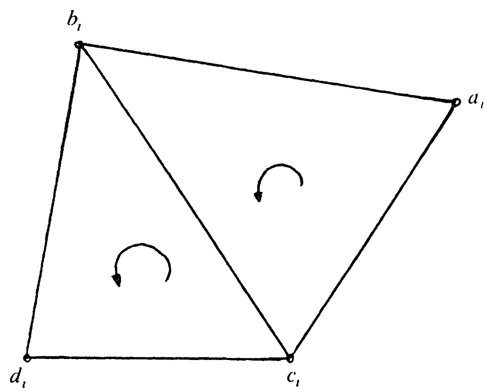


FIG. 2

Hence $\partial(\Sigma_{i=1}^{D} d_i) = s' - s'' + I_1 + I_2$, where

$$I_1 = \sum_{i=1}^{0} [(b_i, d_i) - (a_i, c_i)] = 0$$

by the assumption $\partial s' = 0$ and

$$I_2 = \sum_{i=1}^{0} \left[(d_i, c_i) + (c_i, d_i) \right].$$

But $(x, y) + (y, x) = \partial(y, x, y) + \partial(y, y, y)$, so I_2 is a boundary and s' and s'' are homologous. By induction one concludes that s' and

$$\sum_{i=1}^{D} n_i \left(\left(1 - \frac{k}{H}\right) a_i, \left(1 - \frac{k}{H}\right) b_i \right)$$

are homologous for any $1 \leqslant k \leqslant H$; particularly when $k = H$ one gets that s' is homologous to 0. (See Fig. 2.)

Now, let us assume that $r: B^2 \to S^1$ is a retraction and $i: S^1 \to B^2$ is the inclusion map. If s is the cycle from the statement (a), one has $s = r_\#(i_\#(s))$. From (b) it follows that $i_\#(s)$, and therefore $r_\#(i_\#(s))$, is a boundary which contradicts (a).

To prove the theorem in the general case it is more convenient to replace the pair (B^n, S^{n-1}) by the homeomorphic pair $(\Delta_n, \dot{\Delta}_n)$, where Δ_n is an n-dimensional simplex with $0 \in \operatorname{int} \Delta_n$ and $\dot{\Delta}_n$ its boundary. The proof is now completely analogous to the case $n = 2$, except that the role of θ is played by the function $\theta(a_0, a_1, \ldots, a_{n-1}) = \det[a_0, \ldots, a_{n-1}]$ while $s \in C_{n-1}(\dot{\Delta}_n)$ is obtained in the natural way (see Fig. 3).

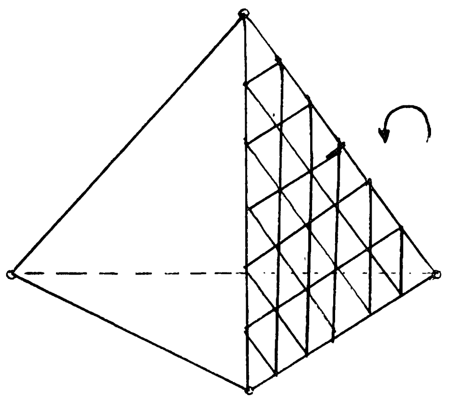


FIG. 3

References

1. S. Eilenberg and N. Steenrod, Foundation of Algebraic Topology, Princeton University Press, Princeton, NJ, 1952.
2. S. Garavaglia, Homology with equationally compact coefficients, Fund. Math., 100 (1978) 89–95.
3. H. J. Keisler, Elementary Calculus, Prindle, Weber and Schmidt, Boston, 1976.
4. ______, Student Supplement to accompany Elementary Calculus, Prindle, Weber and Schmidt, Boston, 1976.
5. ______, Foundation of Infinitesimal Calculus, Prindle, Weber and Schmidt, Boston, 1976.
6. R. Lutz and M. Goze, Nonstandard Analysis, Lecture Notes in Mathematics 881, Springer-Verlag, 1981.
7. M. C. McCord, Non-standard analysis and homology, Fund. Math., 74 (1972) 22–28.
8. J. P. Reveilles, Infinitésimaux et Topologie, Publ. I.R.M.A., Strasbourg, 1983.
9. A. Robinson, Non-standard Analysis, North-Holland, Amsterdam, 1966.
10. E. H. Spanier, Algebraic Topology, McGraw-Hill, New York, 1966.
11. K. D. Stroyan and W. A. J. Luxemburg, Introduction to the Theory of Infinitesimals, Academic Press, New York, 1976.
12. F. Wattenberg, Nonstandard analysis and the theory of shape, Fund. Math., 98 (1978) 41–60.

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

IS THERE AN ALL-PURPOSE TILE?*

BRANKO GRÜNBAUM
Department of Mathematics, University of Washington, Seattle, WA 98195

G. C. SHEPHARD
University of East Anglia, University Plain, Norwich NR4 7TJ, England

The seventeen wallpaper groups (or plane crystallographic groups, see Coxeter & Moser [**1**], Fejes Tóth [**2**] or Schattschneider [3] for descriptions and detailed information) can be exhibited conveniently as the symmetry groups of geometric objects such as tilings (tessellations). Our problem is this: *is there a single tile of which congruent copies can be used to tile the plane in* 17 *different ways, so that the symmetry groups of these tilings are precisely the* 17 *wallpaper groups?*

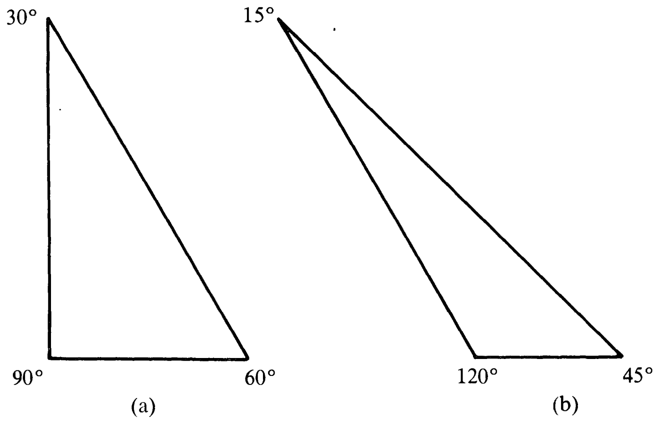


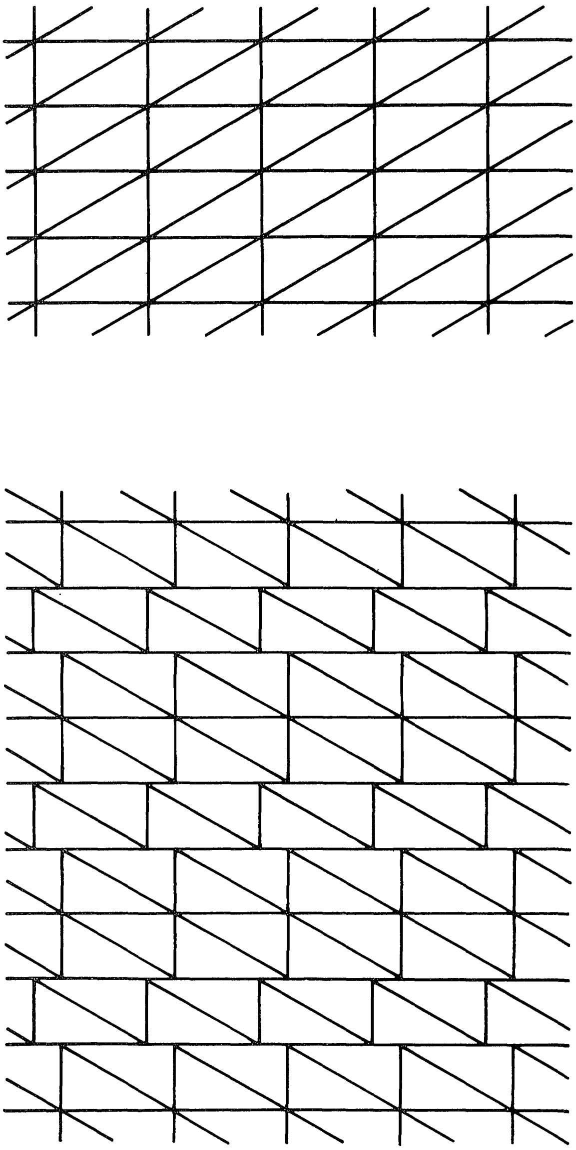
FIG. 1

The tile shown in Fig. 1(a)—a triangle with angles of 30°, 60°, 90°—can be used in tilings of the plane whose symmetry groups are (at least) 14 of the 17 groups (see Fig. 2). It seems impossible to find such tilings for the remaining three groups: $p4$, $p4m$ and $p4g$. However, this becomes possible if we allow tiles of a second shape, shown in Fig. 1(b), as well—see Fig. 3. But is the admission of a second shape always necessary if all 17 groups are to be displayed?

The tile shown in Fig. 1(a) admits tilings with 4-fold rotational symmetry, but it seems that no tiling with this tile has more than one center of 4-fold symmetry, and this is the reason why it cannot be used for groups $p4$, $p4m$ and $p4g$.

We now formulate several conjectures that elaborate on the first question posed above. It is to be understood that every tile is a closed (convex or nonconvex) topological disk, and that in any tilings the set of tiles covers the plane without gaps or overlaps (that is, the interiors of the tiles are pairwise disjoint).

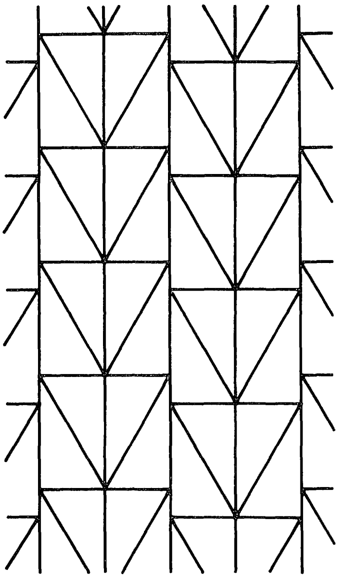
*Research supported by the National Science Foundation grant MCS8301971.



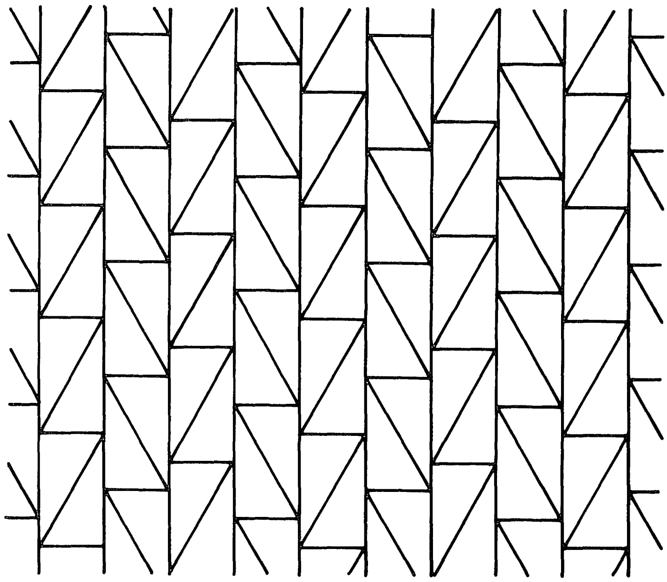
$p2$

$p1$

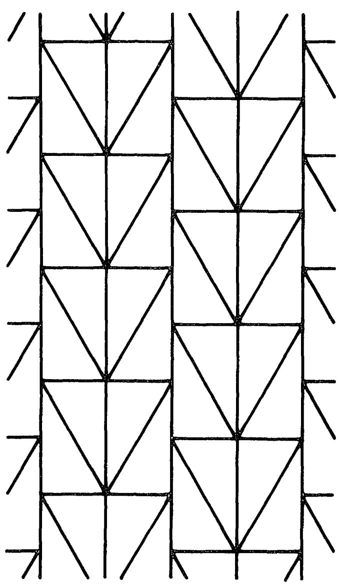
FIG. 2 (first part).



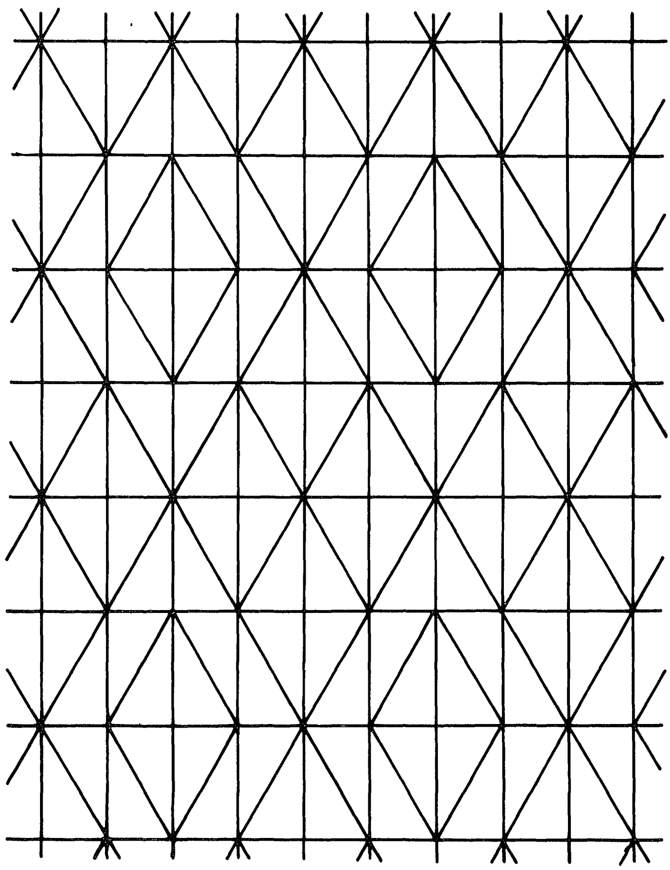
pm



pg

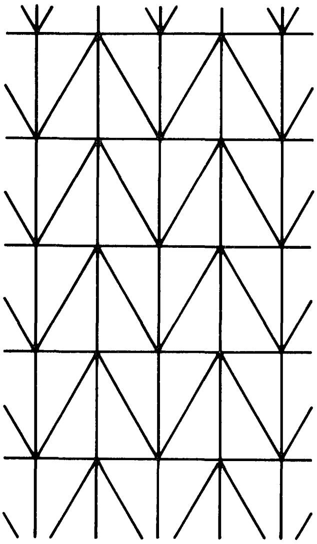


cm

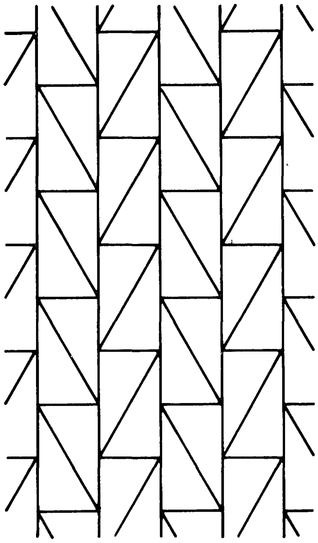


pmm

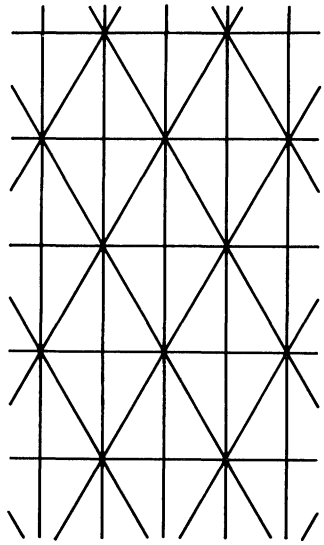
FIG. 2 (second part).



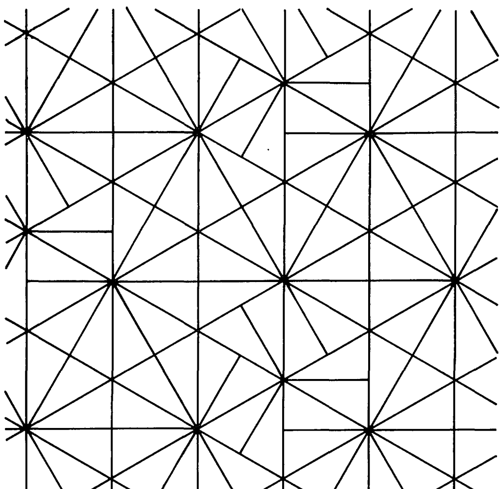
pmg



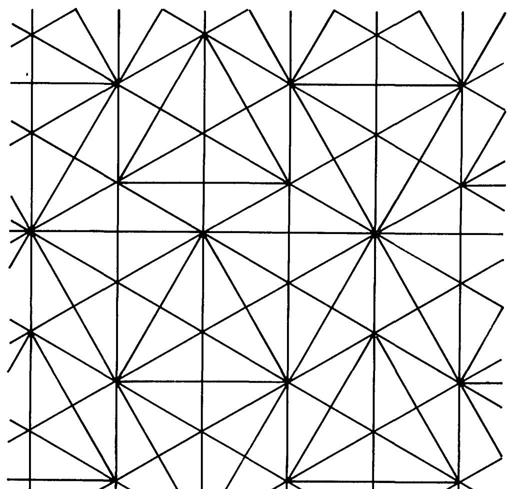
pgg



cmm

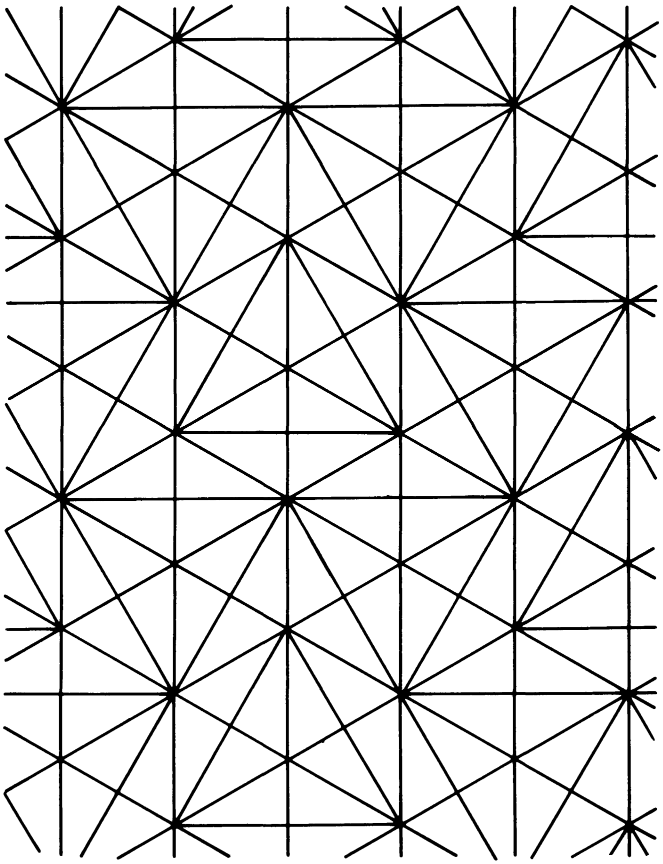


*p*3

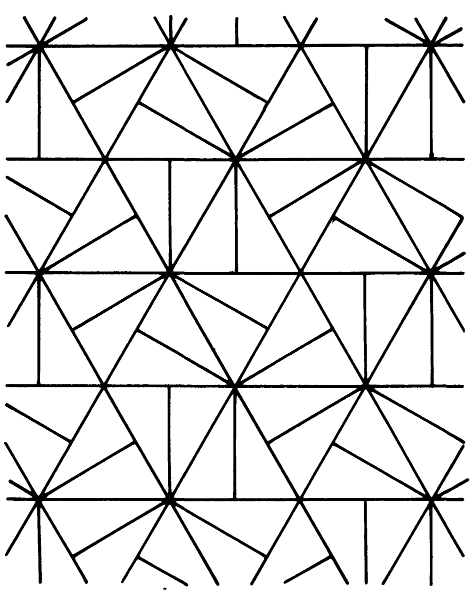


*p*3*m*1

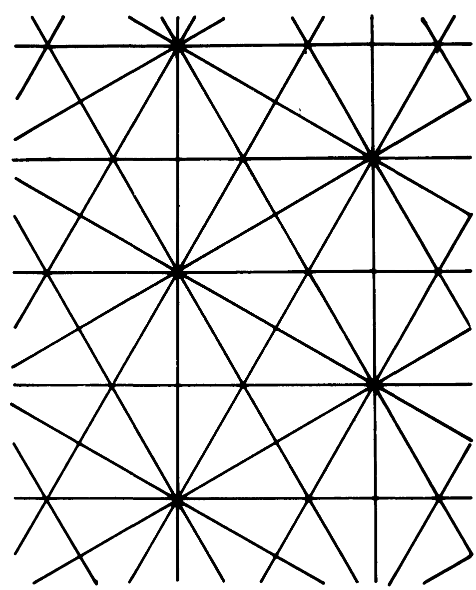
FIG. 2 (third part).



$p31m$



$p6$



$p6m$

FIG. 2 (fourth part).

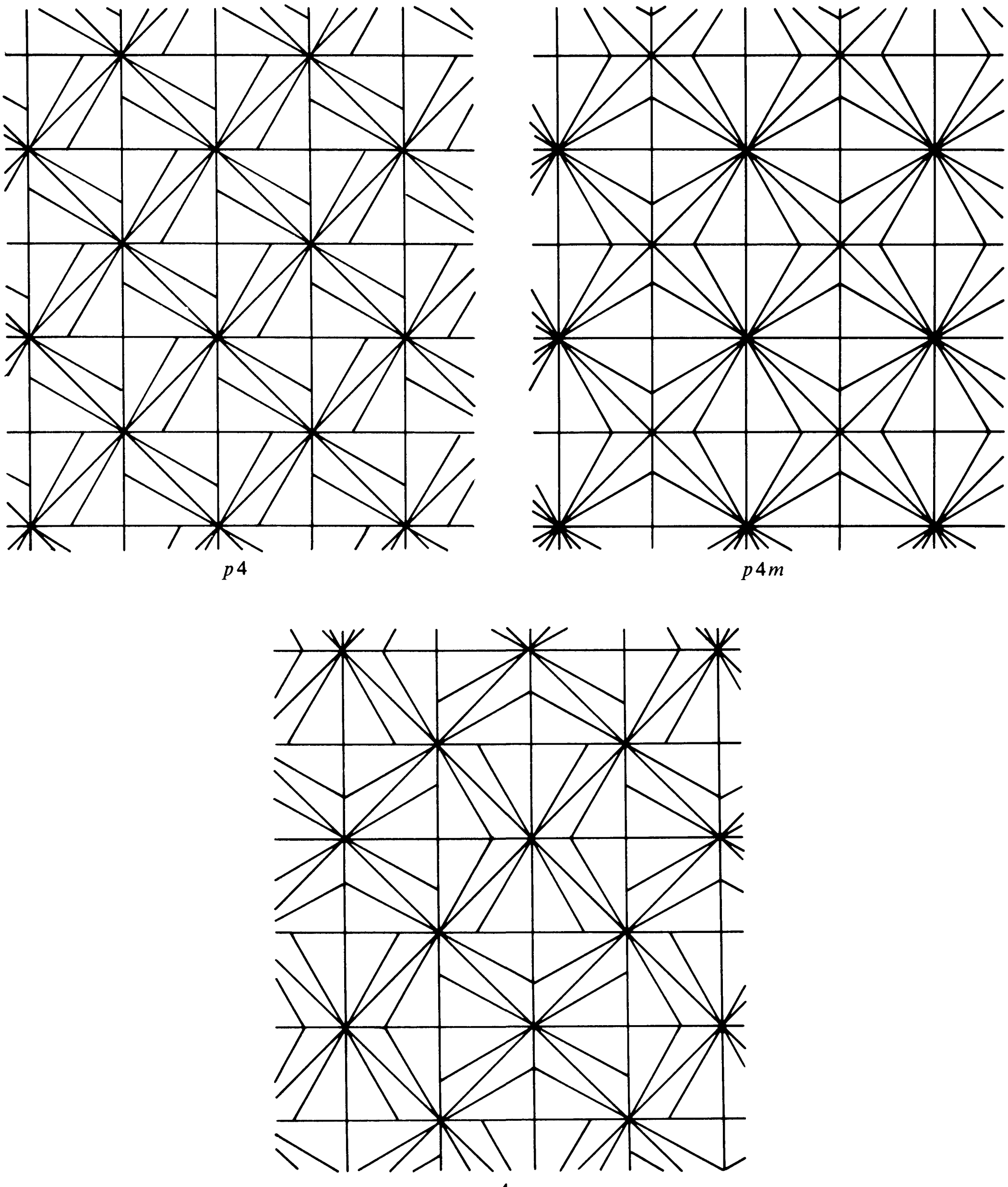


FIG. 3

CONJECTURE 1. *The tile of Fig.* 1(a) *is the only one which leads to tilings with as many as* 14 *of the* 17 *wallpaper groups as symmetry groups.*

CONJECTURE 2. *The tile of Fig.* 1(a) *cannot be used in a tiling with more than one center of* 4*-fold rotational symmetry.*

CONJECTURE 3. *There exists no tile of which congruent copies can be used in two tilings, such that the first has at least two centers of* 3*-fold rotational symmetry and the second has at least two centers of* 4*-fold rotational symmetry.*

References

1. H. S. M. Coxeter and W. O. J. Moser, Generators and Relations for Discrete Groups, 4th ed., Springer, Berlin, 1980.

2. L. Fejes Tóth, Reguläre Figuren, Akadémiai Kiadó, Budapest, 1965. English translation: Regular Figures, Pergamon, New York, 1964.

3. D. Schattschneider, The plane symmetry groups: their recognition and notation, this MONTHLY, 85(1978) 439–450.

NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.
For instructions about submitting Notes for publication in this department see the inside front cover.

FORMS OF THE RESULTANT OF TWO POLYNOMIALS

S. BARNETT
School of Mathematical Science, University of Bradford, Bradford BD7 1DP, England

A resultant is a scalar function of two polynomials which is nonzero if and only if the polynomials are relatively prime. The theory of resultants is an old and much-studied topic in what used to be called the theory of equations, and it is therefore not surprising that a recently proposed form of resultant [4] is not, in fact, new. Indeed, excellent recent expositions including this result have been given by Householder in [5] and the text [6].

The basic ideas are as follows. Consider three formal power series

$$f(\lambda) = a_0 + a_1\lambda + a_2\lambda^2 + \cdots, \quad a_0 \neq 0,$$

$$g(\lambda) = b_0 + b_1\lambda + b_2\lambda^2 + \cdots,$$

and

$$h(\lambda) = c_0 + c_1\lambda + c_2\lambda^2 + \cdots$$

where

$$h(\lambda) = g(\lambda)/f(\lambda).$$

By comparing coefficients in the identity $g(\lambda) = f(\lambda)h(\lambda)$ we see that

$$a_0c_k + a_1c_{k-1} + \cdots + a_kc_0 = b_k, \quad k = 1,2,\dots. \tag{1}$$

Construct a Hankel matrix $H = [h_{ij}]$ of order n with $h_{ij} = c_{m-n+i+j-1}$, where $c_i = 0$ for $i < 0$, and a *bigradient*, or *Sylvester-type*, matrix of order $m + n$:

$$S = \begin{bmatrix} a_0 & a_1 & a_2 & & \cdots & a_n & 0 & \cdots & 0 \\ 0 & a_0 & a_1 & & \cdots & a_{n-1} & a_n & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdots & \cdot & \cdot & \cdot & \cdot \\ 0 & b_0 & b_1 & & \cdots & b_m & 0 & \cdots & 0 \\ b_0 & b_1 & b_2 & \cdots & b_m & 0 & \cdot & \cdots & 0 \end{bmatrix}, \tag{2}$$

where there are m rows of a's and n rows of b's. Then it is straightforward to show by writing the expressions (1) in matrix form, and applying a generalization of Cramer's rule, that

$$\det S = a_0^{m+n}\det H. \tag{3}$$

C E N T E R   S E C T I O N
(Vol. 93, No. 7, August-September 1986)

Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T: Textbook	P: Professional Reading	1-4: Semesters
C: Computer Software	L: Undergraduate Library	** : Special Emphasis
S: Supplementary Reading	13: Grade Level	?? : Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the Monthly.

General, S, L. $E = mc^2$: Picture Book of Relativity. Gerald Kahan. TAB Books, 1983, x + 118 pp, $10.25 (P). [ISBN: 0-8306-0180-5] A non-rigorous exposition of relativity for the non-scientist, developed from several years of popular talks on the topic. The ideas of relativity, excluding the difficult concept of simultaneity, are described in readable text and numerous helpful illustrations; historical background and photographs included. RB

Precalculus, T(13: 1), C, S. MacAlgebra: BASIC Algebra on the Macintosh. Marvin and Rebecca Marcus, Catherine Baczynski. Computer Science Pr, xviii + 427 pp, $24.95 (P); Student and Teacher Diskette, $20 each. [ISBN: 0-88175-135-9] An integrated text on topics from college algebra (sets, numbers, functions, combinatorics, probability) that simultaneously introduces Microsoft Basic on the Macintosh. Standard hand exercises are mixed with computer discovery exercises based on 25 programs available on the accompanying disk, and with other exercises whose programmed solutions are found on the disk. Programs are simple, routine, and not internally documented (often there is no hint in the code of what the program is supposed to do), although the text provides excellent line-by-line discussion of all interesting features. Requires a Microsoft Basic disk and a standard 128K Macintosh. Thin on details of college algebra, but rich in potential for exploration and discovery learning. LAS

Finite Mathematics, T(13-15: 1), S*, L*. The Art of Decision-Making. Morton Davis. Springer-Verlag, 1986, viii + 92 pp, $19.95. [ISBN: 0-387-96228-X] A delightful collection of everyday decision problems, each designed to expose common sense as an unreliable guide, and to indicate effective problem-solving techniques as a substitute. The mathematics includes probability and decision trees, game theory, and the mathematics of voting. The style is informal and interactive; no exercises LCL

History, S(13-16), L. Charles Babbage: On the Principles and Development of the Calculator. Ed: Philip and Emily Morrison. Dover, 1961, xxxviii + 400 pp, $7.95 (P). [ISBN: 0-486-24691-4] A collection of papers by Babbage on such topics as analytic engines, difference engines, railways, Cambridge, picking locks, and travel tips. Included are contemporary reviews of Babbage's calculating machines by Lardner and Menabrea. SG

Logic, T(15-16: 1), L. Algebraic Logic. S.G. Gindikin. Transl: Robert H. Silverman. Prob. Books in Math. Springer-Verlag, 1985, xviii + 356 pp, $46. [ISBN: 0-387-96179-8] Translation of a Russian text for undergraduate logic course published in 1972. Most of the book is devoted to logical function approach to classical propositional logic. Brief coverage of networks of functional elements, probabilistic logic, predicate logic, multi-valued logics. Each chapter contains explanatory text, problems, hints, solutions. KS

Foundations, P. Lecture Notes in Mathematics-1106: Techniques of Admissible Recursion Theory. C.T. Chong. Springer-Verlag, 1984, vii + 214 pp, $11 (P). [ISBN: 0-387-13902-8] Monograph describes techniques developed to generalize recursion theory to admissible ordinals. Illustrates techniques by proving results which led to their introduction. Assumes familiarity with classical recursion theory. KS

Discrete Mathematics, P, L*. Report of Committee on Discrete Mathematics in the First Two Years. MAA, 1986, iv + 104 pp, (P). Final report of an ad hoc national committee convened to make recommendations on the role of discrete mathematics. In brief, they recommend a one-year discrete mathematics course as part of the regular freshman-sophomore curriculum, taught by mathematicians, focussed on proof, recursion, induction, modelling, and algorithmic thinking. The report includes detailed course objectives and sample problems, an annotated list of discrete mathematics textbooks, relations to various computer science curriculum recommendations, and a bibliography of related curriculum reports. LAS

Number Theory, T(18), P. Elliptic Functions. K. Chandrasekharan. Grund. der math. Wissenschaften, B. 281. Springer-Verlag, 1985, xi + 189 pp, $48. [ISBN: 0-387-15295-4] A brief exposition covering the basics of the theory. Topics include Weierstrass's elliptic function, zeta and theta functions, the J-function, Dedekind's z-function, and several applications to number theory. Many interesting historical notes are included. SG

Number Theory, P. Euclidean Rings with Two Infinite Primes. F.J. van der Linden. CWI Tract, V. 15. Math Centrum, 1985, vii + 200 pp, Dfl. 27,40 (P). [ISBN: 90-6196-286-2] The author characterizes the rings in the title; they are found either in certain quadratic, cubic, or quartic extensions of the rationals, or in certain algebraic function fields. SG

Number Theory, P. Arithmetic Moduli of Elliptic Curves. Nicholas M. Katz, Barry Mazur. Annals of Math. Stud., No. 108. Princeton U Pr, 1985, xiv + 514 pp, $60; $22.50 (P). [ISBN: 0-691-08349-5; 0-691-08352-5] Gives an account of the arithmetic theory of the moduli spaces of elliptic curves--emphasizing the behavior at primes dividing the "level" both "self-contained and general." LN

Number Theory, P, L. The Riemann Zeta-Function: The Theory of the Riemann Zeta-Function with Applications. Aleksander Ivić. Wiley, 1985, xvi + 517 pp, $49.95. [ISBN: 0-471-80634-X] A monumental survey of recent developments. Major topics include: Voronoi summation formula, approximate functional equations, power moments, zeros, zero-density estimates, and applications to the distribution of primes and divisor problems. Extensive notes and comments included. A valuable contribution to the literature. SG

Number Theory, T(18), P. Introduction to the Construction of Class Fields. Harvey Cohn. Stud. in Adv. Math., V. 6. Cambridge U Pr, 1985, x + 213 pp, $44.50. [ISBN: 0-521-24762-4] An introduction to explicit class field theory over the rationals and imaginary quadratic fields. The final goal is the generation of Abelian extensions of $Q(\sqrt{d})$ by modular functions. Along the way historical background, algebraic preliminaries, density theorems, and genus theory are discussed. SG

Linear Algebra, T(17-18: 1), P. Error-Free Polynomial Matrix Computations. E.V. Krishnamurthy. Texts & Mono. in Comp. Sci. Springer-Verlag, 1985, xv + 154 pp, $32. [ISBN: 0-387-96146-1] On the manipulation of matrices where elements are polynomials over the field of rationals or the ring of integers. Evaluation, inversion and interpolation using "error-free" discrete Fourier transforms and Hensel Codes. Includes a review of needed abstract algebra. RWN

Linear Algebra, P. Linear Algebra and Its Role in Systems Theory. Ed: Richard A. Brualdi, et al. Contemp. Math., V. 47. AMS, 1984, xiii + 506 pp, $36 (P). [ISBN: 0-8218-5041-5] The proceedings of a summer research conference which was held at Bowdoin College from July 29 to August 4, 1984. CEC

Group Theory, P. Lecture Notes in Mathematics-1098: Groups--Korea 1983. Ed: A.C. Kim, B.H. Neumann. Springer-Verlag, 1984, viii + 183 pp, $9.50 (P). [ISBN: 0-387-13890-0] Proceedings of the international conference on combinatorial group theory held in Korea, 1983. The papers, primarily by Western authors, encompass a broad range of topics within this field. RB

Algebra, P. Lecture Notes in Mathematics-1142: Orders and their Applications. Ed: I. Reiner, K.W. Roggenkamp. Springer-Verlag, 1985, x + 306 pp, $20.50 (P). [ISBN: 0-387-15674-7] Proceedings of a 1984 Oberwolfach conference. A collection of 19 papers covering a variety of topics including Galois modules, class groups, representations, zeta functions, and primality testing. SG

Algebra, P. Group Actions on Rings. Ed: Susan Montgomery. Contemp. Math., V. 43. AMS, 1985, x + 277 pp, $27 (P). [ISBN: 0-8218-5046-6] Proceedings of a 1984 conference held at Bowdoin College. The 18 papers cover a broad range of algebraic topics including algebraic groups, Hopf algebras, invariant theory, K-theory, and C*-algebras. SG

Algebra, T(17), S, P, L. $SL_2(R)$. Serge Lang. Grad. Texts in Math., V. 105. Springer-Verlag, 1985, xiv + 428 pp, $39. [ISBN: 0-387-96198-4] This reprint of a 1975 book (TR, August-September 1975) is an exposition of the infinite dimensional representations of $SL_2(R)$ and of the group modulo a discrete subgroup. CEC

Algebra, P. Syzygies. E. Graham Evans, Phillip Griffith. London Math. Soc. Lect. Note Ser., V. 106. Cambridge U Pr, 1985, 124 pp, $15.95 (P). [ISBN: 0-521-31411-9] An exposition of results on syzygies of finite projective dimension over local rings. Suitable for a one-semester course following a basic commutative algebra course. SG

Algebra, T(18), S, P. Free Rings and Their Relations, Second Edition. P.M. Cohn. London Math. Soc. Mono., V. 19. Academic Pr, 1985, xxii + 588 pp, $80. [ISBN: 0-12-179152-1] A substantially revised version of the author's 1971 First Edition (TR, June-July 1972; Extended Review, May 1973); suitable as a second-year graduate text or as an important reference for algebraists. The book is devoted to arithmetic and model theoretic properties of free and semi-free ideal rings. Many exercises to extensive references included. SG

Numerical Analysis, P. Computational Methods in Bifurcation Theory and Dissipative Structures. M. Kubícek, M. Marek. Ser. in Computat. Physics. Springer-Verlag, 1983, xi + 243 pp, $42. [ISBN: 0-387-12070-X] Numerical techniques for solving systems in which the number and nature of the solutions depend upon parameters. Considers both lumped and distributed parameter systems. RWN

Analysis, P. Theory of Nonlinear Age-Dependent Population Dynamics. G.F. Webb. Pure & Appl. Math., V. 89. Dekker, 1985, vi + 294 pp, $78. [ISBN: 0-8247-7290-3] "The purpose of this monograph is to provide a mathematically complete treatment for a general class of models of [continuous, deterministic], nonlinear age-dependent population dynamics." Rigorous, emphasizing theory; presumes integration theory, linear operator theory, functional analysis, and the theory of semigroups of linear operators in Banach spaces. RB

Algebraic Geometry, P*. Rational Points. Gerd Faltings, et al. Aspects of Math., V. E6. Friedr. Vieweg & Sohn (US Distr: Heyden & Son), 1984, 268 pp, $20 (P). [ISBN: 3-528-08593-2] Notes from the 1983/84 seminar at the Max-Plank-Institut in Bonn on Faltings' recent proof of the famous Mordell conjecture. After a descriptive introduction to moduli spaces and heights, the exposition closely follows the original proof. Additional arithmetic surface results. Of great interest to mathematicians concerned with arithmetic algebraic geometry. RB

Algebraic Geometry, P. Riemann-Roch Algebra. William Fulton, Serge Lang. Grund. der math. Wissenschaften, B. 277. Springer-Verlag, 1985, x + 203 pp, $48. [ISBN: 0-387-96086-4] An exposition of the "elementary algebra" that underlies theorems of the Riemann-Roch type. Topics include λ-rings, Chern classes, Grothendieck's R-R theorems, and intersection formulas. SG

Algebraic Geometry, T(18: 2). Introduction to Commutative Algebra and Algebraic Geometry. Ernst Kunz. Transl: Michael Ackerman. Birkhauser Boston, 1985, x + 238 pp, $29.95. [ISBN: 3-7643-3065-1] Mumford in his preface says: "Has filled a long-standing need for an introduction to commutative algebra and algebraic geometry that emphasizes the concrete elementary nature of the objects with which both subjects began." The author in the Forward says he leads "up to some recent results...concerning the representation of algebraic varieties as intersections of the least possible number of hypersurfaces and...with the most economical generation of ideals in Noetherian rings." Includes many exercises. LN

Differential Geometry, P. Lecture Notes in Mathematics-1090: Differential Geometry of Submanifolds. Ed: K. Kenmotsu. Springer-Verlag, 1984, vi + 132 pp, $7.50 (P). [ISBN: 0-387-13873-0] Proceedings of the Kyoto University conference on differential geometry of submanifolds, 1984. "The aims of the conference [were] to present recent researches of young Japanese mathematicians which treated topics mentioned in the title and to stimulate discussions for future studies." Twelve papers. RB

Differential Geometry, P. Tight and Taut Immersions of Manifolds. T.E. Cecil, P.J. Ryan. Research Notes in Math., V. 107. Pitman, 1985, 336 pp, $22.95 (P). [ISBN: 0-273-08631-6] A relatively complete and self-contained survey of recent work in the areas of tight (minimal total absolute curvature) and taut immersions of manifolds into Euclidean space, and the related field of isoparametric hypersurfaces in spheres. Prerequisites: one year graduate course in differential geometry, elements of homology theory, and critical point theory. RB

Algebraic Topology, T(17: 1), S, P, L. Knots. Gerhard Burde, Heiner Zieschang. Stud. in Math., V. 5. Walter de Gruyter, 1985, xii + 399 pp, DM 138. [ISBN: 0-89925-014-9] This book presents the classical theory of knots in 3-space. It starts with an elementary foundation of the theory and from there passes to the standard notions and notations of algebraic topology. Includes exercises and an impressive bibliography. CEC

Algebraic Topology, P. Combinatorial Methods in Topology and Algebraic Geometry. Ed: John R. Harper, Richard Mandelbaum. Contemp. Math., V. 44. AMS, 1985, xviii + 349 pp, $32 (P). [ISBN: 0-8218-5039-3] A survey of recent accomplishments and further directions for research in topology and algebraic geometry using combinatorial methods; proceedings of the Rochester conference, 1982. Topological and combinatorial group theory, knot theory, 3-manifolds, homotopy theory and infinite dimensional topology, 4-manifolds and algebraic surfaces. RB

Algebraic Topology, P. Intersection Cohomology. A. Borel, et al. Progress in Math., V. 50. Birkhauser Boston, 1984, x + 238 pp, $19.95. [ISBN: 0-8176-3274-3] Notes from a 1983 seminar on intersection homology après Goresky/MacPherson at the University of Bern, Switzerland. Piecewise-linear version of intersection homology; sheaf theoretic approach to intersection cohomology; applications. Presumes some familiarity with algebraic topology and sheaf theory. Comprehensive bibliography. RB

Algebraic Topology, T(18: 1), S, P. Cohomology of Sheaves. Birger Iversen. Lect. Notes Ser., No. 55. Aarhus U, 1984, 237 pp, (P). An exposition of cohomology of sheaves over locally compact spaces. Homological algebra (via category theory), sheaf theory, cohomology with compact support, Poincaré duality, local cohomology, characteristic classes, homology, applications to algebraic geometry, sheaves on paracompact spaces. Presumes graduate-level introductions to algebra, algebraic topology, perhaps manifolds; no exercises. RB

Algebraic Topology, P. Algebraic and Differential Topology--Global Differential Geometry. Ed: George M. Rassias. Teubner-Texte zur Math., B. 70. B.G. Teubner, 1984, 348 pp, 36M (P). Papers contributed to commemorate the 90th anniversary of Marston Morse's birth. Works on cobordisms and concordances, the eta invariant, surface groups, knots, K-theory and finite fields, universal bundles over Grassmannians, and a historical note on the discovery of fission. RB

Topology, T(17-18: 1), S, P. General Topology and Homotopy Theory. I.M. James. Springer-Verlag, 1984, 248 pp, $38. [ISBN: 0-387-90970-2] A self-contained textbook based on the author's Oxford course of graduate lectures. The language of category theory, selected point-set topology, base spaces, topological groups, homotopy, covering spaces, fibrations and cofibrations, classification

of G-fiber bundles, absolute retracts and ANR's. Presumes sophistication such as basic point-set theory and group theory. RB

Topology, T(15-17: 1, 2), S, P, L. Beginner's Course in Topology: Geometric Chapters. D.B. Fuks, V.A. Rokhlin. Transl: A. Iacob. Universitext. Springer-Verlag, 1984, xi + 519 pp, $32 (P). [ISBN: 0-387-13577-4] Translation of a Soviet textbook by recognized authors in "elementary" topology, i.e., sophisticated algebraic structures play a subordinate role. Chapters: general topology and homotopies, CW-complexes, smooth manifolds, bundles, homotopy groups. Departures from standard nomenclature. Ambitious for most first undergraduate courses; could make a nice second course in topology. Exercises. RB

Topology, P. Aspects of Topology: In Memory of Hugh Dowker 1912-1982. Ed: I.M. James, E.H. Kronheimer. London Math. Soc. Lect. Note Ser., V. 93. Cambridge U Pr, 1985, xvii + 335 pp, $29.95 (P). [ISBN: 0-521-27815-5] A memorial volume of papers on topics related to or arising from C.H. Dowker's work. Includes survey of knot tabulations by Thistlethwaite and a posthumous paper with Strauss on frames, plus contributions by several well-known topologists, primarily in general topology. Bibliography; obituary from Bulletin of the London Mathematical Society. RB

Topology, P*. Four-Manifold Theory. Ed: Cameron Gordon, Robion Kirby. Contemp. Math., V. 35. AMS, 1984, vii + 528 pp, (P). [ISBN: 0-8218-5033-4] Proceedings of the remarkable AMS conference on four-manifolds at Durham, New Hampshire, July 4-10, 1982. Discussions of Freedman's solution of the four-dimensional Poincaré conjecture and Donaldson's work on smooth four-manifolds with definite intersection form were the foremost planned activities; unplannable was Quinn's completion of the proof of the annulus conjecture at the conference. Discussion of (remaining!) open problems. RB

Operations Research, S(16-17), P. Stochastic Programming with Multiple Objective Functions. I.M. Stancu-Minasian. Transl: Victor Giurgiutin. Math. & Its Applic. D Reidel, 1984, xv + 334 pp, $59. [ISBN: 90-277-1714-1] Linear programming has long been a standard tool in operations research, but there are objections to it: in applications, at least some coefficients are random, and it is unrealistic to only consider one objective function. This book presents state-of-the-art attempts to address these objections. MT

Operations Research, S(17-18), P. Resource Management Concepts for Large Systems. Rajan Suri. Intern. Ser. in Modern Appl. Math. & Comp. Sci., V. 3. Pergamon Pr, 1981, xiii + 83 pp, $14.50. [ISBN: 0-08-026473-5] An exposition of the author's theory of decentralized solutions for resource management problems in large-scale systems. Application to "real-world" examples: a large warehousing system; file allocation in computer networks. Aimed at students, researchers, practitioners in operations research, systems theory; presumes some familiarity with linear algebra and basics of non-linear programming. RB

Optimization. American Journal of Mathematical and Management Sciences. Ed: Edward J. Dudewicz. American Sciences Pr, 1984, 197 pp, $49.75 (P). [ISBN: 0-935950-08-7] Special issue entitled "Statistics and Optimization: The Interface." Articles outline and analyze algorithms for optimization problems including the traveling salesman problem. MT

Dynamical Systems, P. Lecture Notes in Mathematics-1115: Ergodic Theory and Statistical Mechanics. Jean Moulin Ollagnier. Springer-Verlag, 1985, vi + 147 pp, $12 (P). [ISBN: 0-387-15192-3] A text arising from a course at the University of Paris (given by the author and Didier Pinchon) on topological and measure-theoretic dynamical systems, particularly the symbolic dynamical systems of statistical mechanics. "[F]or the study of dynamical systems amenability is the crucial property of the acting group." RB

Dynamical Systems, P. Dynamical Systems and Cellular Automata. Ed: J. Demongeot, E. Golès, M. Tchuente. Academic Pr, 1985, xv + 399 pp, $39.50. [ISBN: 0-12-209060-8] A collection of 36 papers on the applications of discrete and continuous dynamical systems theory to the study of physical systems, biological systems (e.g., plant development, the brain), and computer science. SG

Probability, P. Lecture Notes in Mathematics-1110: Strong Limit Theorems in Non-Commutative Probability. Ryszard Jajte. Springer-Verlag, 1985, vi + 152 pp, $12 (P). [ISBN: 0-387-13915-X] Presents recent extensions of fundamental pointwise convergence theorems in probability theory and ergodic theory to von Neumann algebras. Intended primarily for probabilists, with a pre-requisite of functional analysis. MT

Probability, S(18), P. Lectures on Stochastic Processes. K. Ito. Springer-Verlag, 1984, iii + 233 pp, $7.10 (P). [ISBN: 0-387-12873-5] Measure-theoretic discussion of stochastic processes. Most discussion based on ideas of Markov processes. Not a text; there are no exercises. MT

Probability, T(14-15: 1, 2). An Introduction to Probability Theory with Statistical Applications. Michael A. Golberg. Math. Conc. & Methods in Sci. & Eng., V. 29. Plenum Pr, 1984, xi + 662 pp, $69.50. [ISBN: 0-306-41645-X] Assumes calculus through multiple integration; self-contained, quite complete treatment of probability theory. Begins with set theory, counting probability, conditional probability and independence. Discrete random variables, their distributions, joint distributions, and expectations are all covered before continuous random variables are mentioned. Although a probability text, 75 pages discuss statistics from point estimation through hypothesis testing and regression. MT

Probability, T(18: 1). Probability Theory with Applications. M.M. Rao. Prob. & Math. Stat. Academic Pr, 1984, xii + 495 pp, $49.50. [ISBN: 0-12-580480-6] Heavily dependent on real analysis, text reviews measurability results and the Lebesgue integral in Chapter 1. Succeeding chapters introduce standard topics of probability theory. Applications follow and include estimation, stopping times, martingales, and stochastic processes. MT

Probability, T*(16-17: 1), S, L. Elements of Applied Stochastic Processes, Second Edition. U. Narayan Bhat. Wiley, 1984, xvi + 685 pp, $44.95. [ISBN: 0-471-87826-X] Chapters 1-9 treat theory of Markov, renewal, and stationary processes. Chapters 10-21 deal with applications including queues, information systems, reliability theory, business management, and time series. Major revision of the First Edition (TR, January 1973). MT

Probability, T(14-15: 1), S, L. Probability Theory, An Introduction. Delmar Crabill. U Pr of America, 1983, x + 294 pp, $12.75 (P). [ISBN: 0-8191-3332-9] Assumes calculus; presents most standard introductory material in standard order, but omits distribution of functions of random variables. Clearly written with many examples and exercises. MT

Statistics, S(18), P*. Nonparametric Methods in General Linear Models. Madan Lal Puri, Pranab Kumar Sen. Wiley, 1985, ix + 399 pp, $49.95. [ISBN: 0-471-70227-7] Assumes knowledge of parametric linear models and graduate-level probability theory. First half concentrates on distribution theory of rank order statistics, and the second half illustrates use in statistical inference related to general linear models. MT

Statistics, S(14-18), P. Statistical Adjustment of Data. W. Edwards Deming. Dover, 1964, x + 261 pp, $6.95 (P). [ISBN: 0-486-64685-8] A republication of Deming's 1943 book. Least squares, estimation, curve fitting. FLW

Statistics, P. Calculation of Special Functions: The Gamma Function, the Exponential Integrals and Error-Like Functions. C.G. van der Laan, N.M. Temme. CWI Tract, V. 10. Math Centrum, 1984, iv + 231 pp, Dfl. 33.30 (P). [ISBN: 90-6196-277-3] For each group of functions considered, gives relevant definitions and analytic properties, provides algorithms, implementations, error analysis, and references to tabulated coefficients. Also provides annotated introduction to literature on computation of special functions. MT

Statistics, S(16-18), P, L. Computer-Aided Data Analysis: A Practical Guide. William R. Green. Wiley, 1985, xv + 268 pp, $29.95. [ISBN: 0-471-80928-4] This book shows the programming novice what to look for and how to apply various computer data analysis programs that are useful in the applied sciences. Proceeds from basic techniques (e.g., frequency distributions, two-dimensional scatterplots, computer plotting and computer graphics) to more advanced statistical methods (e.g., multiple regression), and analysis methods for array-oriented data. LCL

Statistics, T(13-14: 1, 2). Statistical Techniques in Business and Economics, Sixth Edition. Robert D. Mason. Richard D Irwin, 1986, xxxii + 985 pp, $33.95. [ISBN: 0-256-03383-8] This new edition (Fourth Edition, TR, November 1978; Fifth Edition, TR, November 1982) includes hypergeometric distributions, the paired difference test, cluster sampling, testing individual regression coefficients, and some rearrangement of topics. FLW

Statistics, P*. Introduction to Variance Estimation. Kirk M. Wolter. Ser. in Stat. Springer-Verlag, 1985, xi + 427 pp, $48. [ISBN: 0-387-96119-4] Gives variance estimation techniques for large complex sample surveys, and demonstrates their use. These techniques consider issues such as nonresponse, measurement errors, and cost. MT

Statistics, C*.** MACSPIN Graphical Data Analysis Software. Andrew W. and David L. Donoho, Miriam Gasko. D^2 Software (PO Box 9546, Austin, TX 78766), 1985, 185 pp, (P). A stunningly powerful tool for visualizing multivariable data by displaying a rotating scatterplot of three variables and animating (through time) along a fourth dimension. Based on earlier PRIM systems (for Projection, Rotation, Isolation, Masking) that run on mainframe computers, MacSpin takes optimal advantage of the Macintosh 68000 CPU, its system software, and user interface. MacSpin on a 512K Macintosh can handle data sets of 500-2000 points (it will run on 128K, but more slowly) and is smoothly and intelligently linked to the Macintosh system for importing data, saving scatterplots, and using the Mac as a terminal for a data set located on a main computer. Includes six sample data sets (earthquakes, random numbers, iris petals, etc.). The 180-page user manual is exceptionally clear and thorough, including not only details on using MacSpin but also statistical, historical and computer system background on the role of visual multivariable data analysis. LAS

Computer Literacy, T*(13), S, L*. Computers and Data Processing. Harvey M. and Barbara Deitel. Academic Pr, 1985, xxiv + 638 pp, $30.38. [ISBN: 0-12-209020-9] A richly illustrated introductory computer literacy text, readily attracting the attention of the uninitiated, co-authored by the author of a leading operating systems textbook. Organization: introduction, hardware, software, computers in business, computers in society. Chapter on structured programming, 80-page appendix on BASIC provide optional programming component. RB

Computer Programming, S(15-17), P, L. Screen Design Strategies for Computer-Assisted Instruction. Jesse M. Heines. Digital Pr, 1984, xiii + 159 pp, $28. [ISBN: 0-932376-28-2] An introduction to design concepts for CAI video displays. Not a how-to-book, nor a tutorial for existing CAI systems, nor a primer on computer graphics, but a basic text for "courseware" producers seeking to create effective instructional materials. Presumes computer literacy and basic educational psychology or

educational technology background. RB

Computer Programming, T(14-16: 1), L. C Made Easy. Herbert Schildt. Osborne McGraw-Hill, 1985, x + 292 pp, $18.95 (P). [ISBN: 0-07-881178-3] A nicely written, thorough coverage of the language. Emphasizes the portable nature of the language, although the intuitive framework is that of MS-DOS (or PC-DOS) on a microcomputer. However, the author, who specifically mentions using Aztec C, skims over a few annoying problems which are unknown in the UNIX framework. For example, no mention is made of the use of and need for the function agetc(), his comments on pointers and integers are ignored a few pages later and do not apply to any C compiler using large memory models for the 8086, and argv[0] is not the program name in MS-DOS. JAS

Software Systems, S(17-18), P. Lecture Notes in Computer Science-187: Time Series Package (TSPACK). Francois S. Chaghaghi. Springer-Verlag, 1985, 305 pp, $18.70 (P). [ISBN: 0-387-15202-4] User's manual for TSPACK, a library of FORTRAN 77 functions designed at the University of Lausanne 1980-83. Includes general directions, calling sequences, and source code for all procedures. MT

Software Systems, P, L**.** The TeXbook. Donald E. Knuth. Addison-Wesley, 1986, ix + 483 pp. [ISBN: 0-201-13447-0] Sixth printing of the TeX bible: a witty, richly-detailed manual for Knuth's powerful typesetting system. This definitive edition detailing the "final" version of TeX is the first of five volumes by the author on computers and typesetting. This volume explains not only how to use TeX but also how it works internally. Worth reading for its own sake, even if one is not a TeX user. (First printing, TR, October 1984.) LAS

Computer Science, T(15-16), S, P, L. The Logic of Programming. Eric C.R. Hehner. Intern. Ser. in Comp. Sci. Prentice-Hall, 1984, 361 pp, $34.95. [ISBN: 0-13-539966-1] An interesting, well-organized approach to rigor and programming. Background logic, types, names, grammars, semantics, programs, data structure, and execution. Develops and uses a language named "Pro." RWN

Computer Science, S(15-17), P. Lecture Notes in Computer Science-183: The Munich Project CIP, Volume 1: The Wide Spectrum Language CIP-L. CIP Language Group. Springer-Verlag, 1985, xi + 275 pp, $16.60 (P). [ISBN: 0-387-15187-7] The initial description and formal definition of the CIP-L program development language, designed for transformational development of programs from formal specifications by the 18-year project in computer-aided intuition-guided programming at the Technical University of Munich. CIP-L is an "abstract scheme language" which coherently offers multi-level features, defined using transformational semantics. RB

Computer Science, T?(14-15). Computer Organization and Programming with an Emphasis on the Personal Computer, Fourth Edition. C. William Gear. Ser. in Computer Organization & Architecture. McGraw-Hill, 1985, xiii + 414 pp. [ISBN: 0-07-023049-8] Although the general concepts of computer organization are presented, the context of Intel chips and CP/M or PC-DOS is so restrictive as to make the chapter on multi-programming and multiprocessors seem like an anomoly. The author does present a lot of the specifics which will be needed for actual programming, although the book successfully avoids being an assembly programming text. However, either the index is weak (several important ideas appeared in the book but not in the index), or the range of examples of processors and operating systems offers a narrow viewpoint. JAS

Computer Science, T(15-17: 1), P, L. Petri Nets. Wolfgang Reisig. EATCS Mono. on Theor. Comp. Sci., V. 4. Springer-Verlag, 1985, x + 161 pp, $24. [ISBN: 0-387-13723-8] A Petri net consists of two disjoint sets and a relation between them. Nets are used to model procedures and organizations involving regulated flows. This revision and translation of a 1982 German textbook introduces basic concepts and illustrates their application to computer-based information systems. Assumes familiarity with matrices, graphs, relations. KS

Applications, P. Complex Systems--Operational Approaches in Neurobiology, Physics, and Computers. Ed: H. Haken. Ser. in Synergetics, V. 31. Springer-Verlag, 1985, x + 365 pp, $39. [ISBN: 0-387-15923-1] Proceedings of the International Symposium on Synergetics held at Schloss Elmau, Bavaria, May 6-11, 1985. Contributions deal with relationships between microscopic and macroscopic levels in such complex systems as various brain functions, evolutionary processes, order and chaos in physical systems, coordination of motion in biological systems and robots, computers and computing. LCL

Applications (Artificial Intelligence), T(16-17), P. The Computer Modelling of Mathematical Reasoning. Alan Bundy. Academic Pr, 1983, xiv + 322 pp, $35. [ISBN: 0-12-141250-4] An introduction to the subject for undergraduates, graduate students, and practicing mathematicians. After a quick introduction to logic, the author covers several examples of theorem proving programs, concept formulation, and the construction of mathematical models. Examples and exercises included. SG

Applications (Artificial Intelligence), P, L*. Conceptual Structures: Information Processing in Mind and Machine. John F. Sowa. Addison-Wesley, 1984, xiv + 481 pp. [ISBN: 0-201-14472-7] An enticing introduction to knowledge-based systems. The author begins by discussing background material from philosophy, psychology, and artificial intelligence. Conceptual graphs are presented as the primary knowledge representation tool. Applications are given to logic and languages. An appendix describing the necessary mathematical background is included. Many interesting notes and a huge bibliography are included. A fascinating book that carries the reader into philosophy, psychology, linguistics as well as artificial intelligence. SG

Applications (Artificial Intelligence), P. Programming Expert Systems in OPS5: An Introduction to Rule-Based Programming. Lee Brownston, et al. Addison-Wesley, 1985, xviii + 471 pp. [ISBN: 0-201-

10647-7] Relatively complete discussion of OPS-5 syntax, programming environment, organization, implementation, programming techniques and style. Discussion in detail of rule-based tools and packages in general, related expert systems tools, and programming techniques. RM

Applications (Biology), P. Lecture Notes in Biomathematics-60: Population Genetics in Forestry. Ed: H.-R. Gregorius. Springer-Verlag, 1985, vi + 287 pp, $19.60 (P). [ISBN: 0-387-15980-0] Proceedings of a 1985 symposium in Göttingen. The three main topics are tree breeding, mating systems and genetic differentiation within and between populations. Contains very little mathematically. TAV

Applications (Chemistry), P. Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences, Second Edition. C.W. Gardiner. Ser. in Synergetics, V. 13. Springer-Verlag, 1985, xix + 442 pp, $34.50 (P). [ISBN: 0-387-15607-0] Written for theoretical physicists and chemists with a strong mathematical background. The author develops the stochastic calculus and uses approximating diffusion processes to analyse a wide variety of Markov processes. Several amendations and extensions of the First Edition (TR, April 1984). TAV

Applications (Chemistry), S(17-18), P, L. Mathematical Crystallography. M.B. Boisen, Jr., G.V. Gibbs. Reviews in Mineralogy, V. 15. Mineralogical Soc of Amer, 1985, xii + 406 pp, $14 (P). [ISBN: 0-939950-19-7] A self-contained derivation of the 32 crystallographic point groups, the 14 Bravais lattice types, and the 230 crystallographic space group types, written by a mathematician and a mineralogist. The book, accessible and self-teaching, lays the mathematical foundation needed to solve numerous basic problems in crystallography. Numerous worked examples; problem sets. LCL

Applications (Cognitive Science), T(18), P. Pattern Recognition: Human and Mechanical. Satosi Watanabe. Wiley, 1985, xv + 570 pp, $44.95. [ISBN: 0-471-80815-6] An introduction to the subject by one of its major practitioners. Analyzes pattern recognition from several points of view including perception, categorization, induction, entropy minimalization, covariance diagonalization, and statistical decision making. Makes extensive use of linear algebra and mathematical statistics. SG

Applications (Economics), S(18), P. Bioeconomic Modelling and Fisheries Management. Colin W. Clark. Wiley, 1985, xii + 291 pp, $44.95. [ISBN: 0-471-87394-2] A discussion, using mathematical models, of the overfishing problem: to study the relationship between the economic forces affecting the fishing industry (e.g., royalties, taxes, quota allocation systems), and the biological factors that determine the production and supply of fish in the sea. Dynamic programming and optimal control theory, used extensively, is briefly described within the text. LCL

Applications (Economics), S(16-18), P, L. Mathematical Methods in Economics. Ed: Frederick van der Ploeg. Handbook of Applic. Math. 6. Wiley, 1984, xix + 580 pp, $49.95. [ISBN: 0-471-90422-8] Twenty papers written by and for economists, to describe how mathematics is used as a tool in economics. Mathematical prerequisites are referenced to six core volumes that are part of the series. Illustrations from both microeconomics and macroeconomics. LCL

Applications (Economics), P. Game Theory in the Social Sciences, Volume 2: A Game-Theoretic Approach to Political Economy. Martin Shubik. MIT Pr, 1984, viii + 744 pp, $47.50. [ISBN: 0-262-19219-5] In this volume Shubik applies game-theoretic methods developed in the first volume to unify several branches of political economy. Extensive bibliography. KS

Applications (Economics), P. Lecture Notes in Economics and Mathematical Systems-253: Economic Theory and International Trade in Natural Exhaustible Resources. Cees Withagen. Springer-Verlag, 1985, vi + 172 pp, $14.80 (P). [ISBN: 0-387-15970-3] This study is concerned with the economic theory of exhaustible resources (e.g., optimal depletion rate, and pricing). Emphasis on aspects of international trade, which heretofore represents a largely unexplored field of research. The mathematics is that of dynamic programming and optimal control. LCL

Applications (Engineering), P. An Algorithmic Analysis of a Communication Model with Retransmission of Flawed Messages. D.M. Lucantoni. Research Notes in Math., No. 81. Pitman, 1983, 169 pp, $18.95 (P). [ISBN: 0-273-08571-9] A solution to the problem of retransmission after receiving messages containing errors. Problem description, mathematical models, algorithms, and examples. RWN

Applications (Engineering), S(17-18), P. Spectral Techniques and Fault Detection. Ed: Mark G. Karpovsky. Notes & Reports in Comp. Sci. & Appl. Math., V. 11. Academic Pr, 1985, ix + 608 pp, $44.50. [ISBN: 0-12-400060-6] A self-contained presentation of spectral techniques (generalizing Laplace and Fourier transforms) for analysis, design and testing of digital devices; and error detection and correction in such devices. Proceedings of an international workshop (Boston, 1983) intended for graduate students and practicing engineers working in the areas of logic design, spectral techniques and testing/self-testing of digital devices. MU

Applications (Engineering), P. Finite Element Systems: A Handbook. Ed: C.A. Brebbia. Springer-Verlag, 1985, 767 pp, $85. [ISBN: 0-387-15116-8] Third Edition including 47 papers describing the more popular, commercially-available finite element systems. Includes comparisons, applications and use in education. RWN

Applications (Information Theory), T(17: 2), S, P. Elements of Digital Satellite Communication: Channel Coding and Integrated Services, Digital Satellite, Networks, Volume II. William W. Wu. Computer Science Pr, 1985, xiii + 642 pp, $49.95. [ISBN: 0-88175-000-X] As in Volume I (TR, March 1985), the author continues to attempt to identify, explore and analyze selected subjects that are pertinent to either present or future digital satellite communication. Includes chapters on appli-

cations of combinatorial sets in satellite communication, cryptology and message security, and error codes implementation. Includes exercises and references. CEC

Applications (Management), T(17), P. Systems: Concepts, Methodologies, and Applications. Brian Wilson. Wiley, 1984, xvi + 339 pp, $36.95. [ISBN: 0-471-90443-0] A discussion of the "systems" approach to problem solving and consulting especially in the area of management. The author illustrates the methodology with several examples and a case study. Several aspects of model building are discussed. SG

Applications (Medicine), P. Lecture Notes in Medical Informatics-27: Textbank Systems. Erhard Mergenthaler. Springer-Verlag, 1985, vi + 177 pp, $15.50 (P). [ISBN: 0-387-15974-6] A description of the use of a database system in psychoanalytic research. SG

Applications (Physics), S(16-18), P. Open Questions in Quantum Physics. Ed: Gino Tarozzi, Alwyn van der Merwe. Fund. Theories of Physics. D Reidel, 1985, x + 426 pp, $59. [ISBN: 90-277-1853-9] Proceedings of a workshop held in 1983 in Bari, Italy, the aim of which was to discuss the formal, the physical, and the epistemological difficulties of quantum theory in the light of recent crucial developments and to propose some possible resolutions. The papers range in tone from quite technical with numerous mathematical expressions to expository. MU

Applications (Physics), P. Lecture Notes in Physics-239: Geometric Aspects of the Einstein Equations and Integrable Systems. Ed: R. Martini. Springer-Verlag, 1985, 344 pp, $20.50 (P). [ISBN: 0-387-16039-6] Contains the notes of lectures delivered at the Sixth Scheveningen Conference, August 1984. Focused on the area of geometric methods in mathematical physics and particularly on mathematical aspects of the Einstein equations of general relativity and mathematical soliton theory. MU

Applications (Physics), S(16-18), P*. Foundations of the Probabilistic Mechanics of Discrete Media. D.R. Axelrad. Found. of Philo. of Sci. & Technol. Ser. Pergamon Pr, 1984, viii + 166 pp, $25. [ISBN: 0-08-025234-6] Describes and applies probabilistic functional analysis to the mechanics of discrete micro-structures in solids and fluids. Includes the needed background on measure spaces, probability and Markov random fields. MU

Applications (Physics), S(18), P. Heat Conduction Within Linear Thermoelasticity. William Alan Day. Tracts in Nat. Philo., V. 30. Springer-Verlag, 1985, viii + 82 pp, $28. [ISBN: 0-387-96156-9] This tract demonstrates that insight into thermomechanical interaction can be obtained by analyzing the one-dimensional version of the equations of linear thermoelasticity for a homogeneous and isotropic body. These equations constitute the simplest generalization of Fourier's heat equation by incorporating the effects of thermomechanical coupling and inertia. Lucidly written in the style of a graduate-level text in pure mathematics. MU

Applications (Physics), T(13-14), S, L. The Theory of Classical Dynamics. J.B. Griffiths. Cambridge U Pr, 1985, xiv + 315 pp, $39.50. [ISBN: 0-521-23760-2] Devoted to the mature classical theory, the copious discussion places this theory within the context of modern physics and explicates all aspects with remarkable clarity and thoroughness. While utilizing calculus throughout to solve many interesting problems, the text is unusual for the space devoted to discussion rather than computation. A high-quality liberal arts rather than engineering text. MU

Applications (Physics), P. Lecture Notes in Mathematics-1148: Probability Distributions in Quantum Statistical Mechanics. Mark A. Kon. Springer-Verlag, 1985, v + 120 pp, $9.80 (P). [ISBN: 0-387-15690-9] Written to provide a rigorous mathematical foundation for the study of the probability distributions of observables in quantum statistical mechanics, and to apply the theory to examples of physical interest. Covers value functions, integrals of independent random variables, singular central limit theorems, the Lebesgue integral, integrability criteria, and joint distributions. MU

Applications (Social Science), P. Abstract Measurement Theory. Louis Narens. MIT Pr, 1985, viii + 334 pp, $40. [ISBN: 0-262-14037-3] Measurement theory deals with the creation of scales to represent empirical or qualitative structures mathematically. This book presents recent developments including classification of possible scale types, use of nonstandard analysis in the non-Archimedean situations, and application of model theory to study axiomatizability of structures. KS

Reviewers

DFA: David F. Appleyard, Carleton; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; RD: Roger Day, St. Olaf; JD-B: John Dyer-Bennet, Carleton; GF: Giovanna Fjelstad, St. Olaf; SG: Steven Galovich, Carleton; BH: Bruce Hanson, St. Olaf; PH: Paul Humke, St. Olaf; KK: Kenneth Kaminsky, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; AM: Alan Magnuson, St. Olaf; SM: Steve McKelvey, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MT: Michael Tveite, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

References

1. H. S. M. Coxeter and W. O. J. Moser, Generators and Relations for Discrete Groups, 4th ed., Springer, Berlin, 1980.
2. L. Fejes Tóth, Reguläre Figuren, Akadémiai Kiadó, Budapest, 1965. English translation: Regular Figures, Pergamon, New York, 1964.
3. D. Schattschneider, The plane symmetry groups: their recognition and notation, this MONTHLY, 85(1978) 439–450.

NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.

For instructions about submitting Notes for publication in this department see the inside front cover.

FORMS OF THE RESULTANT OF TWO POLYNOMIALS

S. BARNETT

School of Mathematical Science, University of Bradford, Bradford BD7 1DP, England

A resultant is a scalar function of two polynomials which is nonzero if and only if the polynomials are relatively prime. The theory of resultants is an old and much-studied topic in what used to be called the theory of equations, and it is therefore not surprising that a recently proposed form of resultant [4] is not, in fact, new. Indeed, excellent recent expositions including this result have been given by Householder in [5] and the text [6].

The basic ideas are as follows. Consider three formal power series

$$f(\lambda) = a_0 + a_1\lambda + a_2\lambda^2 + \cdots, \quad a_0 \neq 0,$$

$$g(\lambda) = b_0 + b_1\lambda + b_2\lambda^2 + \cdots,$$

and

$$h(\lambda) = c_0 + c_1\lambda + c_2\lambda^2 + \cdots$$

where

$$h(\lambda) = g(\lambda)/f(\lambda).$$

By comparing coefficients in the identity $g(\lambda) = f(\lambda)h(\lambda)$ we see that

$$a_0c_k + a_1c_{k-1} + \cdots + a_kc_0 = b_k, \quad k = 1, 2, \ldots. \tag{1}$$

Construct a Hankel matrix $H = [h_{ij}]$ of order n with $h_{ij} = c_{m-n+i+j-1}$, where $c_i = 0$ for $i < 0$, and a *bigradient*, or *Sylvester-type*, matrix of order $m + n$:

$$S = \begin{bmatrix} a_0 & a_1 & a_2 & & \cdots & a_n & 0 & \cdots & 0 \\ 0 & a_0 & a_1 & & \cdots & a_{n-1} & a_n & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdots & \cdot & \cdot & \cdot & \cdot \\ 0 & b_0 & b_1 & & \cdots & b_m & 0 & \cdots & 0 \\ b_0 & b_1 & b_2 & \cdots & b_m & 0 & \cdot & \cdots & 0 \end{bmatrix}, \tag{2}$$

where there are m rows of a's and n rows of b's. Then it is straightforward to show by writing the expressions (1) in matrix form, and applying a generalization of Cramer's rule, that

$$\det S = a_0^{m+n}\det H. \tag{3}$$

Since det S is a resultant associated with the polynomials

$$a(\lambda) = a_0\lambda^n + a_1\lambda^{n-1} + \cdots + a_n,$$
$$b(\lambda) = b_0\lambda^m + b_1\lambda^{m-1} + \cdots + b_m,$$

it follows immediately that det H is an alternative form of the resultant, and this is just the theorem in [**4**]. Householder's result (3) is somewhat more general than one due to Netto (1895), whose work is summarised in Volume 4 of Muir's *History* [**8**], page 326.

An interesting route to establishing the resultant property of det S is first to introduce a companion matrix A associated with $a(\lambda)$, and use the almost obvious fact that the determinant of the nth order matrix

$$b(A) = b_0A^m + b_1A^{m-1} + \cdots + b_mI$$

is also a resultant for $a(\lambda)$ and $b(\lambda)$. It can then be shown that the matrix in (2) satisfies

$$XS = \begin{bmatrix} Y & Z \\ 0 & b(A)J \end{bmatrix}, \tag{4}$$

where X and Y are triangular matrices, and J is the reverse unit matrix (i.e., the identity matrix with its columns reversed in order). Taking determinants of both sides of (4) produces the desired result relating det S and det $b(A)$. The expression (4) also provides a nice way of linking the Sylvester matrix with Bézout's form of the resultant, which is yet another determinant of order n. Finally, the way is opened to applications in control theory on realizing that $b(A)$ is the controllability matrix for an appropriately defined linear system [**1**].

In conclusion, despite its age, the subject of resultants is an interesting one for the classroom for several reasons: it can be used to illuminate several areas of matrix theory, and to relate to problems on the location of roots of polynomials; moreover, there are many relevant applications in the theory of linear control systems, including important extensions to polynomial matrices. For further details, and references to the original sources, the books [**2**], [**3**], [**7**] should be consulted.

References

1. S. Barnett, A matrix circle in linear control theory, Bull. Inst. Math. Appl., 12 (1976) 173–177.
2. ______, Polynomials and Linear Control Systems, Marcel Dekker, New York, 1983.
3. ______, Matrices in Control Theory, 2nd ed., Krieger, Malabar, Florida, 1984.
4. H. U. Gerber, Wronski's formula and the resultant of two polynomials, this MONTHLY, 91 (1984) 644–646.
5. A. S. Householder, Bigradients and the problem of Routh and Hurwitz, SIAM Rev., 10 (1968) 56–66.
6. ______, The Numerical Treatment of a Single Nonlinear Equation, McGraw-Hill, New York, 1970.
7. T. Kailath, Linear Systems, Prentice-Hall, Englewood Cliffs, NJ, 1980.
8. T. Muir, The Theory of Determinants in the Historical Order of Development, Dover, New York, 1960.

INFINITE BRANCHES OF THE PHI-TREE

ELIOT T. JACOBSON AND ALAN E. PARKS

Department of Mathematics, Ohio University, Athens, OH 45701

Let ϕ denote Euler's totient function. We make the set $\mathscr{I}$ of integers greater than 1 into the vertices of a directed graph, connecting $x \in \mathscr{I}$ with an arrow "pointing down to" $\phi(x)$. For $x \in \mathscr{I}$ with $x \geqslant 3$, the sequence $x, \phi(x), \phi^2(x), \ldots$ eventually reaches 2, hence 2 is the unique minimal element of $\mathscr{I}$. Thus, $\mathscr{I}$ is a tree under this structure, the *Phi-tree*. This idea is due to H. Shapiro [**1**].

For $x, y \in \mathscr{I}$, we say that y is *above* x if there is a directed path from y to x (of course, y is above x if $\phi^k(y) = x$ for some positive integer k). Although the Phi-tree is easily constructed,

one may appreciate the difficulty of predicting for given x which y lie above it by trying to find the 77 elements above $x = 40$, for example. Surprisingly, it is easy to determine which x have infinitely many elements of $\mathscr{J}$ above them, given our main theorem.

An even integer of the form $2^e 3^f$ will be called a 2, 3-*number*.

THEOREM. *The integer $x \geqslant 2$ has infinitely many elements of the Phi-tree above it if and only if x is a 2, 3-number.*

We will give the proof in a sequence of lemmas, beginning with the observations for positive integers e, f that $\phi(2^e 3^f) = 2^e \cdot 3^{f-1}$ and that $\phi(2^e) = 2^{e-1}$. From this the following is clear.

LEMMA 1. *If x is a 2, 3-number, then there are infinitely many elements of the Phi-tree above x. Furthermore, every element below x is a 2, 3-number.*

Of course we have just established one direction of the theorem.

For an integer y define $\nu(y)$ to be the exponent of 2 in the prime factorization of y. The idea of the rest of the proof is to keep track of the sequence $\nu(y), \nu(\phi(y)), \nu(\phi^2(y)), \dots$.

LEMMA 2. *Let $y \in \mathscr{J}$ not be a power of 2. Then $\nu(y) \leqslant \nu(\phi(y))$ with equality implying that $y = 2^e p^f$, where e, f are positive integers and p is a prime with $p \equiv 3(4)$. If, in addition, $\nu(y) = \nu(\phi^2(y))$, and y is not a 2, 3-number, then $f = 1$.*

Proof. Let p be an odd prime divisor of y and write $y = 2^e p^f m$, where $p \nmid m$ and $e = \nu(y)$. Then

$$\begin{aligned}\nu(\phi(y)) &= \nu(\phi(2^e)p^{f-1}(p-1)\phi(m)) = \nu\left(\phi(2^e)p^{f-1} \cdot 2\frac{p-1}{2} \cdot \phi(m)\right)\\ &= \nu(\phi(2^e)) + 1 + \nu\left(\frac{p-1}{2}\right) + \nu(\phi(m)).\end{aligned}$$

But $\nu(\phi(2^e)) \geqslant e - 1$ with equality provided $e \geqslant 1$. Thus

$$\nu(\phi(y)) \geqslant e - 1 + 1 = \nu(y)$$

as claimed. If $\nu(\phi(y)) = \nu(y)$ then $\nu(\phi(2^e)) = e - 1$, so that $e \geqslant 1$. Also $\nu\left(\frac{p-1}{2}\right) = 0$, so that $p \equiv 3(4)$, and finally $\nu(\phi(m)) = 0$ so that $m = 1$, since m is odd. This completes the proof of the first assertion.

For the second assertion, assume that $y = 2^e p^f$, where p is a prime with $p \equiv 3(4)$. If y is not a 2, 3-number, then $p \equiv 3(4)$ implies $p \geqslant 7$ and so $(p-1)/2 \geqslant 3$, whence $2 \mid \phi\left(\frac{p-1}{2}\right)$. But also $(p-1)/2$ is odd, and so in the calculation

$$\phi(y) = 2^e \cdot p^{f-1} \cdot \left(\frac{p-1}{2}\right),$$

the factors 2^e, p^{f-1}, and $(p-1)/2$ are pairwise coprime. Thus

$$\phi^2(y) = \phi(2^e)\phi(p^{f-1})\phi\left(\frac{p-1}{2}\right)$$

and

$$\begin{aligned}\nu(\phi^2(y)) &= e - 1 + \nu(\phi(p^{f-1})) + \nu\left(\phi\left(\frac{p-1}{2}\right)\right)\\ &\geqslant e - 1 + \nu(\phi(p^{f-1})) + 1 \geqslant \nu(y) + \nu(\phi(p^{f-1})).\end{aligned}$$

If now $\nu(\phi^2(y)) = \nu(y)$, then we must have $\nu(\phi(p^{f-1})) = 0$, whence $f = 1$.

The next result, which is fairly well known, we include for the sake of completeness.

LEMMA 3. *For $x \in \mathscr{J}$ the set $\{y | \phi(y) = x\}$ is finite.*

Proof. If $\phi(y) = x$ and if p^e is a prime power divisor of y, then $p^{e-1}(p-1)$ divides x. It follows that $p \leqslant x + 1$ and that

$$e \leqslant \log_p(x) + 1 \leqslant \log_2(x) + 1.$$

A rather crude bound

$$y \leqslant [(x+1)!]^{\log_2(x)+1}$$

follows, and the lemma is thereby proved.

LEMMA 4. *Let $x \in \mathscr{J}$ have infinitely many elements of the Phi-tree above it. Then there is an infinite sequence a_n of integers with $a_1 = x$ and $\phi(a_i) = a_{i-1}$ for all $i \geqslant 2$.*

Proof. Elementary graph theory: Assume for $n \geqslant 1$ that $a_1 = x, a_2, \dots, a_n$ have been constructed with $\phi(a_i) = a_{i-1}$ for $i \geqslant 2$ and with a_n having infinitely many elements of the Phi-tree above it. By Lemma 3 the set $\{y | \phi(y) = a_n\}$ is finite, and so some such y has infinitely many elements above it. Put $a_{n+1} = y$. Continuing in this way we obtain our sequence.

LEMMA 5. *Let $x \in \mathscr{J}$ have infinitely many elements above it. Then x is a 2,3-number.*

Proof. Given such x, construct a sequence a_n as in Lemma 4. If x is not a 2,3-number, then by Lemma 1, none of the a_n is a 2,3-number. By Lemma 2

$$\nu(a_1) \geqslant \nu(a_2) \geqslant \cdots.$$

Thus there is some $n \geqslant 1$ for which

$$\nu(a_n) = \nu(a_{n+k}) \quad \text{for all } k \geqslant 0.$$

For $k \geqslant 2$, we have

$$\nu(a_{n+k}) = \nu(a_{n+k-2}) = \nu(\phi^2(a_{n+k})),$$

and so by Lemma 2,

$$a_{n+k} = 2^{e_k} p_k \quad \text{for } k \geqslant 2,$$

where $e_k \geqslant 1$ and p_k is a prime with $p_k \equiv 3 \pmod 4$. Now

$$2^{e_{k-1}} p_{k-1} = a_{n+k-1} = \phi(a_{n+k}) = 2^{e_k} \frac{p_k - 1}{2},$$

from which we conclude that

$$\frac{p_k - 1}{2} = p_{k-1}.$$

Thus, $p_k = 2p_{k-1} + 1$, and it follows from induction that

$$p_k = 2^{k-2}(p_2 + 1) - 1 \quad \text{for } k \geqslant 2. \tag{1}$$

Since the p_k are odd primes, put $k = p_2 + 1$ in (1) so that

$$p_k = 2^{p_2 - 1}(p_2 + 1) - 1 \equiv 0 \pmod{p_2},$$

so then $p_k = p_2$, a clear contradiction to (1). This completes the proof of the lemma and the theorem.

Reference

1. H. Shapiro, An arithmetic function arising from the ϕ function, this MONTHLY, 50 (1943) 18–30.

HOMOMORPHISMS ON $C(R)$

RICHARD M. ARON
Department of Mathematics, Kent State University, Kent, OH 44240

GERD H. FRICKE
Department of Mathematics, Wright State University, Dayton, OH 45435

We recall that $C(R)$ is the vector space of continuous real-valued functions on the real line R, and that a nonzero mapping $\varphi\colon C(R) \to R$ is a homomorphism if

$$\varphi(af + bg) = a\varphi(f) + b\varphi(g) \quad \text{and} \quad \varphi(fg) = \varphi(f)\varphi(g) \text{ for all } a, b \in R \text{ and } f, g \in C(R).$$

There is no lack of proofs of the result that every homomorphism $\varphi\colon C(R) \to R$ is point evaluation at some point of R. Nevertheless, the following proof, which does not appear to be widely known, may be of interest because of its simplicity and accessibility.

THEOREM. *If $\varphi\colon C(R) \to R$ is a homomorphism, then there is some real number $c \in R$ such that $\varphi(f) = f(c)$ for all $f \in C(R)$.*

Proof. Note that $\varphi(1) = \varphi(1)\varphi(1)$, and since φ is not identically zero we can conclude that $\varphi(1) = 1$; here 1 denotes the constant function as well as the real number. Let $\varphi(t) = c$, where $t \in C(R)$ is the identity function. We show that $\varphi(f) = f(c)$ for all $f \in C(R)$. For $f \in C(R)$, let $k(t) = f(t) - f(c)$. Then $k(c) = 0$ and

$$\varphi(k) = \varphi(f) - \varphi(f(c)) = \varphi(f) - f(c).$$

Thus $\varphi(k) = 0$ if and only if $\varphi(f) = f(c)$, and therefore we only need to show that $\varphi(f) = 0$ whenever $f(c) = 0$.

Suppose first that f is identically 0 in an open interval containing c. Define $g(t)$ by

$$g(c) = 0 \quad \text{and} \quad g(t) = \frac{f(t)}{t - c} \text{ for } t \neq c.$$

Then $g \in C(R)$ and $\varphi(f) = \varphi(g) \cdot \varphi(t - c) = 0$.

Now assume that there is some function $f \in C(R)$ satisfying $f(c) = 0$ but $\varphi(f) \neq 0$. Since φ is linear, we may assume that $\varphi(f) = 1$. Using the continuity of f, we see that there is some interval $(c - \varepsilon, c + \varepsilon)$ about c on which $|f(t)| < 1/2$. Define $h \in C(R)$ by $h(t) = f(t)$ in $(c - \varepsilon, c + \varepsilon)$, $h(t) = f(c + \varepsilon)$ for $t \geqslant c + \varepsilon$, and $h(t) = f(c - \varepsilon)$ for $t \leqslant c - \varepsilon$. Then h and f agree near c, and so by our initial remark, $\varphi(f - h) = 0$. Finally since $(1 - h)^{-1} \in C(R)$, $1 = \varphi((1 - h)^{-1})\varphi(1 - h) = 0$. This contradiction proves that no such f exists, and the theorem is established. Q.E.D.

The proof can be easily extended in several ways. For example, to see that all homomorphisms $\varphi\colon C^k(R) \to R$ are point evaluations, the above proof works, although one must do more work to find a C^k function h on R such that $h = f$ near c and $|h(t)| < 1/2$ for all t.

A NOTE ON KRONECKER'S APPROXIMATION THEOREM

KA-LAM KUEH
Department of Mathematics, Columbia University, New York, NY 10027

There are several proofs [**1**, D1, ..., D9], [**2**], [**3**, Chap. XXIII] of the two versions A and B below of Kronecker's theorem on simultaneous inhomogeneous approximation. The object of this note is to give a simultaneous inductive proof of the two versions. In what follows, $\|\beta\|$ denotes the distance from the real number β to the nearest integer, and $\mathbb{Q}$ denotes the set of rational numbers.

A_n. *If the real numbers $\theta_1, \ldots, \theta_n$ are linearly independent over $\mathbb{Q}$, then for real $\alpha_1, \ldots, \alpha_n$ and $\varepsilon > 0$ there are arbitrarily large real t for which*

$$\|t\theta_i - \alpha_i\| < \varepsilon \qquad (i = 1, \ldots, n). \tag{1}$$

B_n. *If in A_n the numbers $\theta_1, \ldots, \theta_n, 1$ are independent over $\mathbb{Q}$, then in* (1) *t may be taken to be an arbitrarily large integer.*

Proof. A_1: This follows by taking $t = (T + \alpha_1)/\theta_1$ for arbitrarily large integral T.

A_n *implies* B_n: By Dirichlet's theorem (see [3, Theorem 201]), there are integers $m, m_1, \ldots, m_n$ with $m > 0$ for which

$$|m\theta_i - m_i| < \frac{\varepsilon}{2} \quad (i = 1, \ldots, n).$$

Since the $m\theta_i - m_i$ are independent, it follows from A_n that for some arbitrarily large t,

$$\| t(m\theta_i - m_i) - \alpha_i \| < \frac{\varepsilon}{2} \qquad (i = 1, \ldots, n).$$

These two inequalities, along with the triangle inequality for $\| \cdot \|$, give

$$\|[t](m\theta_i - m_i) - \alpha_i\| < \varepsilon \qquad (i = 1, \ldots, n),$$

which is (1) with $T = [t]m$.

B_n *implies* A_{n+1}: Since $\theta_1, \ldots, \theta_{n+1}$ are independent over $\mathbb{Q}$, so are $\theta_1/\theta_{n+1}, \ldots, \theta_n/\theta_{n+1}, 1$. It follows from B_n that for some arbitrarily large integral T,

$$\| T(\theta_i/\theta_{n+1}) - (\alpha_i - \alpha_{n+1}\theta_i/\theta_{n+1}) \| < \varepsilon \qquad (i = 1, \ldots, n),$$

which is (1) with $t = (T + \alpha_{n+1})/\theta_{n+1}$. We also have

$$\|t\theta_{n+1} - \alpha_{n+1}\| = \|T\| = 0,$$

so A_{n+1} holds and the proof is complete.

References

1. H. Bohr, Collected Mathematical Works, Dansk Matematisk Forening, København, vol. III, 1952.
2. T. Estermann, A proof of Kronecker's Theorem by induction, J. London Math. Soc., 8 (1933) 18–20.
3. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, 5th ed., Oxford Univ. Press, 1977.

WHEN IS A POINT x SATISFYING $\nabla f(x) = 0$ A GLOBAL MINIMUM OF f?

J.-B. HIRIART-URRUTY
Laboratoire d'Analyse Numérique, Université Paul Sabatier, 118, route de Narbonne, 31062 Toulouse Cedex, France

When trying to minimize a differentiable function f on $\mathbb{R}^n$, we are led to the following first-order necessary condition:

$$\text{if } f \text{ has a minimum on } \mathbb{R}^n \text{ at } \bar{x}, \quad \text{then } \nabla f(\bar{x}) = 0. \tag{1}$$

This condition is sometimes referred to as Fermat's rule in recognition that Fermat discovered around 1630 a method, which actually amounted to the elements of differential calculus, of drawing tangents to curves and finding maxima and minima. Students making their first steps in Optimization verify easily that Condition (1) does not ensure that $\bar{x}$ is a minimum of f on $\mathbb{R}^n$ and that some extra assumption like *convexity* of f should be added for making this condition

sufficient. In accordance with the terminology used in Optimization, a point $\bar{x}$ satisfying $\nabla f(\bar{x}) = 0$ is called a *stationary* point of f.

At this stage one may ask: *what additional condition should a stationary point of f satisfy for being a minimum of f?* Our aim in this short note is to give a simple and concise answer to this question. For that purpose, we need to recall some elementary facts from Convex Analysis.

1. Basic Results from Convex Analysis. Let $\mathbb{R}^n$ be equipped with the usual inner product denoted by $\langle \cdot , \cdot \rangle$. We consider a function f: $\mathbb{R}^n \to \mathbb{R}$. Given $x^* \in \mathbb{R}^n$, we consider the collection of all affine functions $a(x) = \langle x, x^* \rangle - \alpha$ with slope x^* that minorize f, so

$$(2) \qquad \langle x, x^* \rangle - \alpha \leqslant f(x) \quad \text{for all } x \in \mathbb{R}^n.$$

This collection, if nonempty, gives rise to the smallest α^* for which (2) holds. If there is no affine function with slope x^* minorizing f, we agree to set $\alpha^* = +\infty$. In any case,

$$(3) \qquad \alpha^* = \sup_{x \in \mathbb{R}^n} \{ \langle x, x^* \rangle - f(x) \}.$$

The function that assigns α^* to x^* is precisely what is called the *conjugate function* of f and is denoted by f^*. By reiterating the operation $f \to f^*$ on f^*, we get the *biconjugate* of f, defined for all $x \in \mathbb{R}^n$ by

$$(4) \qquad f^{**}(x) = \sup_{x^* \in \mathbb{R}^n} \{ \langle x, x^* \rangle - f^*(x^*) \}.$$

Of course, if f^* is identically equal to $+\infty$, then f^{**} is identically equal to $-\infty$.

One of the key results in Convex Analysis is that f^{**} is the "convexified version" of f. The following theorem makes this precise.

THEOREM. *f^{**} is the pointwise supremum of the collection of all affine functions on $\mathbb{R}^n$ majorized by f.*

Proof. An affine function a: $x \to a(x) = \langle x, x^* \rangle - \alpha$ is majorized by f if and only if:

$$\alpha \geqslant \sup_{x \in \mathbb{R}^n} \{ \langle x, x^* \rangle - f(x) \} = f^*(x^*).$$

Therefore

$$\sup\{ a(x) : a \text{ is affine and } a \leqslant f \} = \sup_{x^* \in \mathbb{R}^n} \{ \langle x, x^* \rangle - f^*(x^*) \} = f^{**}(x). \text{ Q.E.D.} \ \square$$

Thus f^{**} is a convex function on $\mathbb{R}^n$; it coincides with f if f is convex.

EXAMPLE 1. Let f: $\mathbb{R} \to \mathbb{R}$ be defined by $f(x) = (x^2 - 1)^2$ for all $x \in \mathbb{R}$. Then $f^{**}(x) = [\max(0, x^2 - 1)]^2$ for all $x \in \mathbb{R}$.

EXAMPLE 2. Let f: $\mathbb{R} \to \mathbb{R}$ be defined as:

$$f(x) = |x| \text{ if } |x| \leqslant 1, \quad 2 - |x| \text{ if } 1 \leqslant |x| \leqslant 3/2, \quad |x| - 1 \text{ if } |x| \geqslant 3/2.$$

Then $f^{**}(x) = f(x)$ when $|x| \geqslant 3/2$, while $f^{**}(x) = |x|/3$ whenever $|x| \leqslant 3/2$.

Another basic fact we need to recall concerns differentiation of convex functions. We consider a *convex* function f: $\mathbb{R}^n \to \mathbb{R}$. Given $\bar{x} \in \mathbb{R}^n$ and a direction $d \in \mathbb{R}^n$, the function $\alpha \to [f(\bar{x} + \alpha d) - f(\bar{x})]/\alpha$ is an increasing function of $\alpha > 0$. So

$$\inf_{\alpha > 0} [f(\bar{x} + \alpha d) - f(\bar{x})]/\alpha = \lim_{\alpha \to 0^+} [f(\bar{x} + \alpha d) - f(\bar{x})]/\alpha$$

exists; it is called the directional derivative of f at $\bar{x}$ in the direction d and denoted by $f'(\bar{x}; d)$. The convexity of f implies that $f'(\bar{x}; d)$ is convex as a function of d; in particular,

$$f'(\bar{x}; d_1 + d_2) \leqslant f'(\bar{x}; d_1) + f'(\bar{x}; d_2) \quad \text{for all } d_1 \text{ and } d_2 \text{ in } \mathbb{R}^n.$$

If $f'(\bar{x}; d)$ is linear in d, then f is differentiable at $\bar{x}$; in such a case, $f'(\bar{x}; d) = \langle \nabla f(\bar{x}), d \rangle$.
See [1] for further results about Convex Analysis.

2. A Necessary and Sufficient Condition for Optimality

THEOREM. *Let $f: \mathbb{R}^n \to \mathbb{R}$ be a differentiable function. Then $\bar{x} \in \mathbb{R}^n$ is a minimum of f on $\mathbb{R}^n$ if and only if*

$$\text{(i)} \quad \nabla f(\bar{x}) = 0$$

and

$$\text{(ii)} \quad f^{**}(\bar{x}) = f(\bar{x}).$$

In such a case f^{**} is differentiable at $\bar{x}$ and $\nabla f^{**}(\bar{x}) = 0$.

Drawing some examples of functions on $\mathbb{R}$ makes it easier to understand the meaning of (ii). In Example 1, $\bar{x} = +1$ or -1 is a minimum of f on $\mathbb{R}$, while $\bar{x} = 0$ satisfies (i) but not (ii).

Note that if f is convex, then $f^{**} = f$, so Condition (ii) is always satisfied.

Proof of the theorem. Let $\bar{x}$ be a minimum of f on $\mathbb{R}^n$. The constant function $a: x \to a(x) = f(\bar{x})$ is an affine function majorized by f. Since f^{**} is the pointwise supremum of the collection of all affine functions majorized by f, we get that $f^{**}(\bar{x}) = f(\bar{x})$.

Conversely, let $\bar{x}$ be a stationary point of f satisfying $f^{**}(\bar{x}) = f(\bar{x})$. Since $f^{**}(x) \leqslant f(x)$ for all $x \in \mathbb{R}^n$, we have that:

$$[f^{**}(\bar{x} + \alpha d) - f^{**}(\bar{x})]/\alpha \leqslant [f(\bar{x} + \alpha d) - f(\bar{x})]/\alpha \quad \text{for all } \alpha > 0 \quad \text{and} \quad d \in \mathbb{R}^n.$$

Passing to the limit when $\alpha \to 0^+$ yields

$$(f^{**})'(\bar{x}; d) \leqslant \langle \nabla f(\bar{x}), d \rangle = 0 \quad \text{for all } d. \tag{5}$$

Since $-(f^{**})'(\bar{x}; -d) \leqslant (f^{**})'(\bar{x}; d)$, the inequality above implies that $(f^{**})'(\bar{x}; -d) \geqslant 0$ for all d. Therefore

$$(f^{**})'(\bar{x}; d) = 0 \quad \text{for all } d \in \mathbb{R}^n,$$

which amounts to saying that f^{**} is differentiable at $\bar{x}$ with $\nabla f^{**}(\bar{x}) = 0$. Consequently, $\bar{x}$ is a minimum of the convex function f^{**} and, thereby, of f. □

Reference

1. R. T. Rockafellar, Convex Analysis, Princeton University Press, 1970.

THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

For instructions about submitting material for publication in this department see the inside front cover.

A FIRST DERIVATIVE TEST FOR FUNCTIONS OF SEVERAL VARIABLES

MICHAEL W. BOTSKO
Department of Mathematics, Saint Vincent College, Latrobe, PA 15650

In [1, p. 369] the following Second Derivative Test is given for the purpose of studying the relative extrema of functions of two or more variables.

THEOREM 1. *Let f be a real-valued function having all of its second order partial derivatives*

continuous in D a neighborhood centered at a in R^n, and let a be a critical point of f. Then at a the function f has

(i) *a minimum if* $(\alpha \cdot \nabla)^2 f(a) > 0$ *for every unit vector* α;
(ii) *a maximum if* $(\alpha \cdot \nabla)^2 f(a) < 0$ *for every unit vector* α;
(iii) *a saddle point if* $(\alpha \cdot \nabla)^2 f(a)$ *can change sign.*

(∇ is the del operator in R^n and $(\alpha \cdot \nabla)^2 f(a) = \sum_{i,j} \alpha_i \alpha_j \dfrac{\partial^2 f(a)}{\partial x_j \partial x_i}$, where α_i is the ith component of α.)

In two dimensions the conditions can be rewritten as:

(i) $f_{xx}(a) > 0, f_{xx}(a) f_{yy}(a) - (f_{xy}(a))^2 > 0$ (minimum);
(ii) $f_{xx}(a) < 0, f_{xx}(a) f_{yy}(a) - (f_{xy}(a))^2 > 0$ (maximum);
(iii) $f_{xx}(a) f_{yy}(a) - (f_{xy}(a))^2 < 0$ (saddle point).

To my knowledge no calculus text gives criteria for studying local extrema of functions of several variables using only first derivatives. It is the purpose of this paper to show that such criteria can easily be stated and proved. Our main Theorem follows:

THEOREM 2. *Let f be a real-valued function which is continuous on a neighborhood D centered at a in R^n and differentiable on $D \setminus \{a\}$. Then*

(1) *f has a local maximum value at a if* $(x - a) \cdot \nabla f(x) < 0$ *for all x in $D \setminus \{a\}$;*
and
(2) *f has a local minimum value at a if* $(x - a) \cdot \nabla f(x) > 0$ *for all x in $D \setminus \{a\}$.*

(These conclusions are intuitively clear if we note that $(x - a) \cdot \nabla f(x)$ is simply the product of $\|x - a\|$ and the directional derivative of f at x in the direction $\dfrac{x - a}{\|x - a\|}$, a direction pointing away from a.)

Proof. Let us only prove the first part of the conclusion. Suppose that $(x - a) \cdot \nabla f(x) < 0$ on $D \setminus \{a\}$ and let $x \in D \setminus \{a\}$. Define $F:[0,1] \to R$ such that $F(t) = f(a + t(x - a))$. Clearly F is continuous on $[0,1]$ and differentiable on $(0,1)$ with

$$F'(t) = (x - a) \cdot \nabla f(a + t(x - a)).$$

By the Mean Value Theorem there exists $t_0 \in (0,1)$ such that $F(1) - F(0) = F'(t_0)$, which implies that

$$\begin{aligned} f(x) - f(a) &= (x - a) \cdot \nabla f(a + t_0(x - a)) \\ &= \frac{1}{t_0}(t_0(x - a)) \cdot \nabla f(a + t_0(x - a)). \end{aligned}$$

Now let $y = a + t_0(x - a)$. Since

$$0 < \|y - a\| = t_0 \|x - a\| < \|x - a\| < r,$$

where r is the radius of D, we see that y belongs to $D \setminus \{a\}$. Since $y - a = t_0(x - a)$ the above equation becomes

$$f(x) - f(a) = \frac{1}{t_0}(y - a) \cdot \nabla f(y)$$

and since $\dfrac{1}{t_0}(y - a) \cdot \nabla f(y) < 0$ by assumption, we have that $f(x) < f(a)$. Therefore $f(x) \leqslant f(a)$ for all x in D, which says that f has a local maximum value at a.

As an example let us consider the function

$$f(x, y) = x^2 + 2xy + 3y^2 + 2x + 10y + 9.$$

Let us test the critical point $(1, -2)$ by the method of this paper.

$$\begin{aligned}(x-1, y+2)\cdot\nabla f(x,y) &= (x-1, y+2)\cdot(2x+2y+2, 2x+6y+10)\\ &= 2(x-1)(x+y+1)+2(y+2)(x+3y+5).\end{aligned}$$

Let $u = x-1$ and $v = y+2$; the above expression becomes

$$2(u^2+2uv+3v^2) = 2\big((u+v)^2+2v^2\big) > 0 \quad \text{for all } (u,v) \neq (0,0).$$

Therefore $(x-1, y+2)\cdot\nabla f(x,y) > 0$ for all $(x,y) \neq (1,-2)$, which says that f has a local minimum value at $(1, -2)$.

It is not difficult to show that if a critical point can be classified as either a maximum or a minimum point using Theorem 1, then it can also be classified using Theorem 2. To see this suppose f and a are as in the hypotheses of Theorem 1 and suppose $(\alpha\cdot\nabla)^2 f(a) > 0$ for all unit vectors α. Let

$$m = \min\{(\alpha\cdot\nabla)^2 f(a) : \alpha \text{ is a unit vector}\}.$$

Certainly m exists since $(\alpha\cdot\nabla)^2 f(a)$ is a continuous function of α on the closed, bounded set $\{\alpha : \|\alpha\| = 1\}$ in R^n. Clearly $m > 0$. Since all second order partial derivatives of f are continuous at a and since

$$\left|(\alpha\cdot\nabla)^2 f(x) - (\alpha\cdot\nabla)^2 f(a)\right| \leqslant \sum_{i,j}\left|\frac{\partial^2 f(x)}{\partial x_j\,\partial x_i} - \frac{\partial^2 f(a)}{\partial x_j\,\partial x_i}\right|$$

for all unit vectors α, obviously there exists $\delta > 0$ such that

$$\left|(\alpha\cdot\nabla)^2 f(x) - (\alpha\cdot\nabla)^2 f(a)\right| < \frac{m}{2}$$

if $\|x-a\| < \delta$ and α is any unit vector.
Thus

$$-\frac{m}{2} < (\alpha\cdot\nabla)^2 f(x) - (\alpha\cdot\nabla)^2 f(a) < \frac{m}{2},$$

which implies that

$$(\alpha\cdot\nabla)^2 f(x) > (\alpha\cdot\nabla)^2 f(a) - \frac{m}{2} \geqslant \frac{m}{2} > 0$$

if $\|x-a\| < \delta$ and α is any unit vector. Therefore $(\alpha\cdot\nabla)^2 f(x) > 0$ if $\|x-a\| < \delta$ and α is any unit vector. Now let x be any point such that $0 < \|x-a\| < \delta$. Let $F(t) = f(a + t(x-a))$ for $0 \leqslant t \leqslant 1$. Obviously F is even defined on some open interval containing $[0,1]$. Clearly

$$F'(t) = (x-a)\cdot\nabla f(a+t(x-a))$$

and

$$F''(t) = ((x-a)\cdot\nabla)^2 f(a+t(x-a))$$

for $0 \leqslant t \leqslant 1$. Letting $\alpha = \dfrac{x-a}{\|x-a\|}$ we have

$$F''(t) = \|x-a\|^2(\alpha\cdot\nabla)^2 f(a+t(x-a)) > 0$$

if $0 \leqslant t \leqslant 1$. Therefore $F'(t)$ is increasing, and since $F'(0) = 0$ we have that $F'(t) > 0$ if $0 < t \leqslant 1$. Thus $F'(1) > 0$ which says that $(x-a)\cdot\nabla f(x) > 0$. Therefore $(x-a)\cdot\nabla f(x) > 0$ for $0 < \|x-a\| < \delta$ and the hypothesis given in part 2 of Theorem 2 holds.

Thus we have shown that if Theorem 1 can be used to conclude that a critical point is a local

minimum point, then Theorem 2 can also be used to draw this conclusion. The same thing is of course true for local maximum points.

On the other hand it is easy to give examples of functions for which Theorem 1 fails and yet Theorem 2 yields a conclusion. Two examples are

$$g(x, y) = 1 - x^{2/3} - y^{4/5} \quad \text{and} \quad h(x, y) = 25 + (x - y)^4 + (x - 1)^4.$$

In practice, when Theorem 1 is applicable, it may often be easier to use than Theorem 2. Nevertheless I believe the latter Theorem should be of interest to undergraduate students because it extends the well-known First Derivative Test to functions of several variables and also because it provides a nice exercise with directional derivatives.

Reference

1. Watson Fulks, Advanced Calculus, Wiley, New York, 1978.

N! AND THE ROOT TEST

CHARLES C. MUMMA II
Department of Mathematics, University of Washington, Seattle, WA 98195

I would be willing to bet that 99.98% of all freshman calculus students (perhaps even more!) attempting to determine the convergence of $\sum_{n=1}^{+\infty} \frac{2^n}{n!}$ would *not* use the root test. This is not surprising, considering that most texts do not evaluate $\lim_{n\to+\infty}\sqrt[n]{n!}$. However, armed with the knowledge that this limit is $+\infty$, the root test becomes more versatile and accessible.

LEMMA. $\lim_{n\to+\infty}\sqrt[n]{n!} = +\infty$.

Proof. First notice that $(2n)! \geqslant \prod_{k=n}^{2n} k \geqslant n^{n+1}$. Consequently

$$\sqrt[2n]{(2n)!} \geqslant \sqrt[2n]{n^{n+1}} \geqslant \sqrt{n} \quad \text{and} \quad \sqrt[2n+1]{(2n+1)!} \geqslant \sqrt[2n+1]{n^{n+1}} \geqslant \sqrt{n}.$$

A more accurate analysis of this limit yields a rather interesting result.

LEMMA. $\lim_{n\to+\infty}\dfrac{\sqrt[n]{n!}}{n} = \dfrac{1}{e}$.

Proof. Taking logarithms, we get

$$\ln\left(\frac{\sqrt[n]{n!}}{n}\right) = \frac{1}{n}\ln(n!) - \ln(n) = \frac{1}{n}\sum_{k=1}^{n}\ln(k) - \frac{1}{n}\sum_{k=1}^{n}\ln(n) = \sum_{k=1}^{n}\ln\left(\frac{k}{n}\right)\frac{1}{n}.$$

As $n \to +\infty$, this becomes $\int_0^1 \ln(x)\,dx = -1$.

This result is usually proved using the fact that the root test is stronger than the ratio test, that is,

$$\liminf\frac{a_{n+1}}{a_n} \leqslant \liminf\sqrt[n]{a_n} \leqslant \limsup\sqrt[n]{a_n} \leqslant \limsup\frac{a_{n+1}}{a_n},$$

where $(a_n)_{n\in\mathbb{N}}$ is an arbitrary sequence of positive real numbers. The beauty of the present proof is its simplicity and directness and its use of methods readily available to freshman calculus students.

THE POWER METHOD FOR FINDING EIGENVALUES ON A MICROCOMPUTER

GARETH WILLIAMS
Department of Mathematics and Computer Science, Stetson University, Deland, FL 32720

DONNA WILLIAMS
Department of Mathematical Sciences, Rollins College, Winter Park, FL 32789

Introduction. Computers are becoming more available for classroom use in mathematics and quality mathematics software is beginning to appear. The time is now ripe to look into the ways that these tools can be used to explore mathematics. Microcomputers can be used for much more than merely drill and practice. They can be used to introduce a flavor of discovery into courses. Students can discover results and view theorems in a pedagogical manner that has not been possible until this era. With all the computational power needed at their fingertips (power that was only available to research workers in the past), students can examine many situations quickly and focus on the behavior of methods or models. The computer introduces an element of surprise. Things do not always work the way that they are expected to. This paper discusses such a situation in the application of the power method to compute the dominant eigenvalue and a corresponding eigenvector of a matrix. It illustrates how much is often learned when things "go wrong".

The Power Method [1]. Numerical techniques exist for evaluating certain eigenvalues and eigenvectors of various types of matrices. The power method is a straightforward iterative method that leads to the dominant eigenvalue (if it exists) and a corresponding eigenvector. It is often taught in linear algebra and numerical methods courses.

The dominant eigenvalue is the one with the largest absolute value. We remind the reader of the power method at this time.

Let A be an $n \times n$ matrix having n linearly independent eigenvectors and a dominant eigenvalue λ. Let X_0 be an arbitrarily chosen initial column vector having n components. If X_0 has a nonzero component in the direction of an eigenvector for λ, then the sequence

$$X_1 = AX_0, \quad X_2 = A\hat{X}_1, \quad X_3 = A\hat{X}_2, \quad \ldots, \quad X_k = A\hat{X}_{k-1}, \quad \ldots$$

will approach an eigenvector for λ. Here $\hat{X}_k$ is a normalized form of X_k, obtained by dividing each component of X_k by the absolute value of its largest component.

Furthermore, the sequence

$$\frac{\hat{X}_1 \cdot A\hat{X}_1}{\hat{X}_1 \cdot \hat{X}_1}, \ldots, \frac{\hat{X}_k \cdot A\hat{X}_k}{\hat{X}_k \cdot \hat{X}_k}, \ldots$$

will approach the dominant eigenvalue.

The power method is not the most efficient numerical method for computing eigenvalues and eigenvectors; convergence can be extremely slow. However, it is the easiest to prove and the most commonly taught in introductory courses.

Construction of a Test Case. Let us construct a 3×3 matrix that has known eigenvalues and eigenvectors for testing the power method. A similarity transformation CAC^{-1} performed on a diagonal matrix A will lead to a matrix B having the diagonal elements of A as eigenvalues and having the columns of C as eigenvectors [2].

Let

$$B = CAC^{-1} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 6 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 6 & 5 & -5 \\ 2 & 6 & -2 \\ 2 & 5 & -1 \end{pmatrix}.$$

The eigenvalues of B are thus known to be $1,4,6$ with corresponding eigenvectors $\begin{pmatrix}1\\0\\1\end{pmatrix}, \begin{pmatrix}0\\1\\1\end{pmatrix}, \begin{pmatrix}1\\1\\1\end{pmatrix}$. For convenience we shall henceforth write the eigenvectors as row vectors.

All computation, such as the similarity transformation above, is carried out by students in a linear algebra class using menu driven software [3]. The students do no programming. The similarity transformation is carried out using a matrix inverse program to compute C^{-1} and then using a matrix multiplication program twice to compute AC^{-1} and then $C(AC^{-1})$. All intermediate results in the computation of CAC^{-1} are saved on the computer for use in the next stage.

Let us verify that B does indeed have the above eigenvalues and corresponding eigenvectors. This can be conveniently done using the following multiplication:

$$\underset{\substack{\uparrow\\ B}}{\begin{pmatrix}6 & 5 & -5\\ 2 & 6 & -2\\ 2 & 5 & -1\end{pmatrix}} \underset{\substack{\nwarrow\ \uparrow\ \nearrow\\ \text{eigenvectors}}}{\begin{pmatrix}1 & 0 & 1\\ 0 & 1 & 1\\ 1 & 1 & 1\end{pmatrix}} = \underset{\substack{\nwarrow\ \uparrow\ \nearrow\\ 1,4,\text{ and } 6 \text{ times}\\ \text{the eigenvectors}}}{\begin{pmatrix}1 & 0 & 6\\ 0 & 4 & 6\\ 1 & 4 & 6\end{pmatrix}}.$$

Thus $(1,0,1)$, $(0,1,1)$ and $(1,1,1)$ are eigenvectors of B corresponding to the eigenvalues 1, 4, and 6. B can now be used to test the power method. The method should result in the dominant eigenvalue 6 and the corresponding eigenvector $(1,1,1)$.

Applying the Power Method. Let us use the vector $X = (1,2,3)$ as the initial vector for the power method. This is a popular initial vector with students for iterative methods! The method, using 15 iterations, gives 3.99926723 as the dominant eigenvalue with corresponding eigenvector

$$(2.94530762 \times 10^{-8}, .999999996, .999999996).$$

These are approximations for the second eigenvalue 4 and corresponding eigenvector $(0,1,1)$. This is an opportunity for students to examine the conditions of the power method—to find out why the expected results have not occurred.

There are three assumptions made in the power method. The first is that there are three (in this case) linearly independent eigenvectors, the second that there exists a dominant eigenvalue, the third that the initial vector has a nonzero component in the direction of the dominant eigenvector. The first two conditions are satisfied. B has three linearly independent eigenvectors and a dominant eigenvalue, namely 6. Thus the third condition involving a nonzero component in the direction of a dominant eigenvector must be violated. Let us write $(1,2,3)$ as a linear combination of eigenvectors. We get

$$(1,2,3) = 1(1,0,1) + 2(0,1,1) + 0(1,1,1).$$

The initial vector $(1,2,3)$ does indeed have a zero component in the direction of the dominant eigenvector $(1,1,1)$. The conditions of the method do not hold. We cannot expect convergence to the dominant eigenvalue.

The next step, of course, is to investigate the convergence to the second eigenvalue. Is this to be expected in general if the initial vector has zero component in the direction of the dominant eigenvector? The proof of the power method is straightforward and students can easily see why, on selecting an initial vector having zero component in the direction of the dominant eigenvector, convergence will take place to the second eigenvalue and corresponding eigenvector if the initial vector has a nonzero component in this direction. This observation suggests how the power method can be used to determine further eigenvalues and eigenvectors once the dominant eigenvalue and eigenvector have been found.

Finally, however, if the power method is continued beyond the 15th iteration, divergence from the eigenvalue 4 takes place with gradual convergence towards the dominant eigenvalue 6! For

example after 60 iterations, the method gives 5.54002294, while after 100 iterations it gives 5.99999995. The theory of the power method does not predict such a phenomenon. It predicts convergence to the eigenvalue 4. This is an opportunity to discuss the effect of round-off errors that occur when such methods are executed on computers. At the 16th iteration, X_{16} has a significant nonzero component in the direction of the dominant eigenvector $(1,1,1)$ due to round-off errors on the computer. We are back with all the original conditions of the power method being satisfied. X_{16} is the new initial vector and convergence takes place to the dominant eigenvalue and eigenvector beyond this point. The convergence is extremely slow due to the small component of X_{16} in the direction of the dominant eigenvector.

The geometrical interpretation of the above is of interest. The initial vector $(1,2,3)$ lies in the subspace spanned by the eigenvectors $(1,0,1)$ and $(0,1,1)$. The theory of the power method shows that convergence should take place within this subspace to the dominant eigenvalue and eigenvector of this subspace, namely 4 and $(0,1,1)$. In our case however, round-off errors cause vectors to stray out of this two-dimensional subspace. This causes vectors beyond the 16th to be gradually dragged toward the eigenvector $(1,1,1)$ which lies outside the subspace.

Comments. This example arose by chance in a linear algebra class. An Apple microcomputer was used to carry out the computations using special linear algebra software. The matrix C was introduced as one having a clean inverse. The matrix A and initial guess were supplied by students. The class (and the instructor!) struggled together to interpret the output. The matrix B is now given to students as a regular assignment on the power method. They are asked to interpret the output. It is this type of experimentation that the computer introduces into the teaching of mathematics.

Let us summarize the benefits that have been achieved through this example.

(1) The student has applied the power method in the environment that it is meant to be applied, namely on the computer.

(2) The importance of conditions under which the method holds has been stressed. The conditions are part of the method—a part that is often overlooked.

(3) The student has had the experience of modifying the method to arrive at a generalization.

(4) The effect of round-off errors on computers has been emphasized.

(5) The importance of interpreting and checking computer output has been illustrated. The student could have assumed that 4 was the dominant eigenvalue if the method had been applied to 15 iterations.

(6) The importance of using test cases, with known results, to understand the behaviour of algorithms before entering upon the unknown has been vividly illustrated.

The time is now ripe to start collecting examples such as the above for use in undergraduate classes. These examples can be used to supplement standard course material, adding an element of mathematical exploration. Students now have the opportunity to see dimensions of the field that could not be revealed to previous generations. At the same time it is important to stress that the computer should be used sparingly and naturally. This is a new teaching medium. We are all going to have to discover when and how to use it most effectively.

The authors are interested in starting a collection of such "mathematical experiments" with the idea of making this collection available to the mathematical community. Readers who have discovered such examples are encouraged to contact us.

References

1. Gareth Williams, Linear Algebra with Applications, Allyn and Bacon, Boston, MA, 1984, p. 411.
2. Ibid., p. 361.
3. Gareth Williams and Donna J. Williams, Allyn and Bacon, Linear Algebra Computer Companion, Boston, MA, 1984. (A software package consisting of two diskettes and a manual.)

PROBLEMS AND SOLUTIONS

For instructions about submitting ***proposed*** *problems for publication in this department see the inside front cover. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

Solutions *should be sent to the address given on the inside front cover.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

For instructions about submitting solutions of these Elementary Problems, which should be mailed by January 31, 1987, see the inside front cover. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgement).

E 3159. *Proposed by Emeric Deutsch, Polytechnic Institute of New York.*

Define a sequence (p_n) by $p_0 = 1$ and

$$p_n + \frac{1}{1!}p_{n-1} + \frac{1}{2!}p_{n-2} + \cdots + \frac{1}{(n-1)!}p_1 + \frac{1}{n!}p_0 = 1.$$

Show that (p_n) is convergent and find its limit.

E 3160. *Proposed by L. I. Nicolaescu, University Ai Cuza Iasi, Romania.*

Prove that

$$\frac{(4n)!(4n-2)!}{\{(2n)!\}^4} \leqslant \left|\frac{B_{4n}B_{4n-2}}{B_{2n}^4}\right|, \quad n \geqslant 2,$$

where B_n denotes the nth Bernoulli number.

E 3161. *Proposed by Ira Gessel, Massachusetts Institute of Technology.*

Let $A = (a_{ij})_{i,j\geqslant 0}$ be the infinite lower triangular matrix defined by $a_{ij} = \binom{i}{j}\binom{i}{i-j}$. Let $A^{-1} = (b_{ij})_{i,j\geqslant 0}$. Show that

$$\sum_{i,j=0}^{\infty} b_{ij}\frac{x^i}{i!^2}y^j = f((1+y)x)/f(x),$$

where

$$f(x) = \sum_{n=0}^{\infty}\frac{x^n}{n!^2} = I_0(2\sqrt{x}). \quad (I_0 \text{ is one of the modified Bessel functions.})$$

E 3162. *Proposed by Paul Monsky, Brandeis University.*

A superqueen is a piece that moves on a square board like an ordinary chess queen but is permitted to continue along the extended diagonals. (One may think of the board as a torus with opposite sides next to one another.) A result of Pólya's that has been rediscovered by others from time to time (see, for example, E2698 [1979, 309]) is that N superqueens may be placed on an N by N board with no two attacking one another if and only if N is prime to 6.

(a) Is it possible, for each value of N, to place $N-2$ superqueens on an N by N board with no two attacking one another?

(b) For what values of N can $N-1$ superqueens be so positioned on an N by N board?

E 3163. *Proposed by Clark Kimberling, University of Evansville.*

Consider the sequence $\{\pi_i\}_{i=2}^{+\infty}$ of permutations of the positive integers defined by

$$\pi_i : \begin{pmatrix} ki+1 & ki+2 & \dots & ki+i \\ ki+i & ki+i-1 & \cdots & ki+1 \end{pmatrix}$$

as k runs through all nonnegative integers, i.e., π_i reverses the i-tuples

$$(1\,2\,3\,\dots\,i),(i+1\quad i+2\,\dots\,2i),\dots$$

of consecutive positive integers. Define $A_n = \pi_n\pi_{n-1}\dots\pi_2(1)$. Prove that the sequence

$$\{A_n\}_{n=2}^{+\infty} = \{2,4,5,7,12,\dots\}$$

is strictly increasing.

E 3164. *Proposed by Hüseyin Demir, Middle East Technical University, Ankara, Turkey.*

Let s, t be the lengths of the tangent line segments to an ellipse from an exterior point. Find the extreme values of the ratio s/t.

SOLUTIONS OF ELEMENTARY PROBLEMS

Eigenvectors of a Cross-Diagonal Matrix

E 2936 [1982, 213]. *Proposed by H. Kestelman, University College, London.*

An $n \times n$ complex matrix $A = [a_{ij}]$ is cross-diagonal if $a_{ij} = 0$ whenever $i + j \neq n + 1$. Find the condition that the eigenvectors of A span $\mathbb{C}^n$, the entire n-vector space.

Solution by Charles Lanski, University of Southern California. Setting $a_i = a_{ij}$ for $i + j = n + 1$, the condition is that either a_i and a_{n+1-i} are both nonzero or both zero, for each $1 \leqslant i \leqslant n/2$. To see this, let A represent the transformation T in the standard basis $\{\mathbf{e}_i\}$, so that $T(\mathbf{e}_i) = a_i\mathbf{e}_{n+1-i}$ for each $1 \leqslant i \leqslant [(n+1)/2]$. Clearly, $\mathbb{C}^n$ is the direct sum of the T-invariant subspaces V_i generated by $\{\mathbf{e}_i, \mathbf{e}_{n+1-i}\}$. If the condition holds, then T has two eigenvectors on each two-dimensional V_i, so A is diagonalizable. If the condition fails for some i, then T has only one independent eigenvector in that V_i, so A cannot be diagonalizable.

The condition given is equivalent to A being semi-simple (i.e., the minimal polynomial for A is a product of distinct irreducible factors), and the above proof is valid when A is over any field of characteristic $\neq 2$.

Editorial Note: Donald R. Robinson and J. Zelmanowitz also observed that the solution is the same for A over any field of characteristic $\neq 2$; and Robinson added that when the characteristic is 2, A is diagonalizable if and only if it is already diagonal. Pei Yuan Wu remarked that this is Exercise 14 in Section 58 of *Finite-dimensional Vector Spaces*, P. R. Halmos; Ulrich Abel reported that it appears with proof on page 94 of *Lineare Algebra* II, K. Doerk, B. Huppert, and E. Kroll.

Also solved by U. Abel (West Germany), M. D. Ašić (Yugoslavia), J. L. Brenner, E. L. Christian and B. Mathis (student), E. Deutsch, C. Georghiou (Greece), M. Golomb, C. R. Hampton, M. Josephy (Costa Rica), W. A. Newcomb, K.-C. Ng (Hong Kong), B. Richter (Canada), D. W. Robinson, E. T. Wong, P. Y. Wu (China), J. Zelmanowitz, and the proposer.

Convergence Connection Between $\sum_{n>0} \frac{1}{f(n)}$ and $\sum_{n>0} \frac{f^{-1}(n)}{n^2}$ for Suitable f

E 2944 [1982, 333]. *Proposed by P. Mancevice, Clark University.*

Let $f(x)$ be a positive, continuous, strictly increasing function defined for $x > 0$ such that $f(x) \to \infty$ as $x \to \infty$. Let f^{-1} denote the inverse function. (a) Show that $\Sigma_{n>0} 1/f(n)$ converges if and only if $\Sigma_{n>0} f^{-1}(n)/n^2$ converges. (b) Show that if $\Sigma_{n>0} 1/f(n)$ converges, then

$$\lim_{n\to\infty} n^{-2} \sum_{f(k)\leqslant n} f(k) = 0.$$

Solution to (a) *by S. Gopalsamy, Indian Institute of Technology, New Delhi, India.* We assume without loss of generality that there exists $x > 0$ such that $f(x) = 1$. The series $\Sigma_{n>0} 1/f(n)$ converges if and only if $\int_1^\infty dx/f(x)$ converges, since $1/f$ is monotonic. Similarly, $\Sigma_{n>0} f^{-1}(n)/n^2$ converges if and only if $\int_1^\infty f^{-1}(x)\,dx/x^2$ converges, since, for any $x \in [m, m+1]$,

$$f^{-1}(m+1)/m^2 \geqslant f^{-1}(x)/x^2 \geqslant f^{-1}(m)/(m+1)^2.$$

Further,

$$\int_1^n \frac{1}{f(x)}\,dx = \int_{f(1)}^{f(n)} \frac{1}{y}\,df^{-1},$$

where we make the substitution $y = f(x)$ and take all integrals to be Riemann-Stieltjes integrals. Note that

$$\lim_{n\to\infty} \int_{f(1)}^{f(n)} \frac{1}{y}\,df^{-1} \text{ is finite if and only if } \lim_{n\to\infty} \int_1^n \frac{1}{y}\,df^{-1} \text{ is finite.}$$

Clearly,

$$\int_1^n \frac{1}{y}\,df^{-1} = \left[\frac{1}{y} f^{-1}\right]_1^n - \int_1^n f^{-1}(y)\,d\left(\frac{1}{y}\right)$$
$$= \frac{f^{-1}(n)}{n} - \frac{f^{-1}(1)}{1} + \int_1^n \frac{f^{-1}(y)}{y^2}\,dy.$$

Hence $\int_1^\infty dx/f(x)$ converges if and only if

$$(1) \qquad \lim_{n\to\infty}\left[\frac{f^{-1}(n)}{n} + \int_1^n \frac{f^{-1}(y)}{y^2}\,dy\right] < \infty.$$

Now the convergence of $\Sigma_{n>0} 1/f(n)$ implies (1) which implies the convergence of $\int_1^\infty (f^{-1}(y)/y^2)\,dy$, and so then $\Sigma_{n>0} f^{-1}(n)/n^2$ converges. Conversely, the convergence of $\Sigma_{n>0} f^{-1}(n)/n^2$ implies that both of the limits, $\lim_{n\to\infty} f^{-1}(n)/n$ and $\lim_{n\to\infty} \int_1^n (f^{-1}(y)/y^2)\,dy$, are finite. This implies (1), so then $\int_1^\infty dx/f(x)$ converges and finally $\Sigma_{n>0} 1/f(n)$ converges.

Solution to (b) *by A. Meir, University of Alberta.* Since $f(x)$ is increasing, if $\Sigma_{n>0} 1/f(n)$ converges then $\lim_{x\to+\infty} x/f(x) = 0$. Now, writing $f^{-1}(n) = x_n$, we have

$$n^{-2} \sum_{f(k)\leqslant n} f(k) = n^{-2} \sum_{k\leqslant x_n} f(k) \leqslant n^{-2} x_n \cdot n = x_n/f(x_n),$$

which converges to 0 as $n \to \infty$.

Also solved by D. Broline, Chico Problem Group, J. Deutsch, V. Hernandez, M. Josephy (Costa Rica), W. A. Newcomb, and A. Villani (Italy). Part (a) only also solved by E. Hertz, J. Riley, R. Stong, R. A. Struble, and the proposer.

A Sharpening of Jordan's Inequality

E 2952 [1982, 424]. *Proposed by A. McD. Mercer, University of Guelph, Ontario, Canada.*

Prove that

$$\sin\theta \geqslant \frac{2}{\pi}\theta + \frac{\theta}{12\pi}(\pi^2 - 4\theta^2) \quad \text{in} \quad 0 \leqslant \theta \leqslant \frac{\pi}{2}.$$

Solution I *by Ulrich Abel, University of Giessen, West Germany.* Let $I = [0, \pi/2]$, $b = 1/3\pi$. We are going to find the greatest real number a such that $\sin x \geqslant ax - bx^3 (x \in I)$. In fact $a = \min_{x \in I} f(x)$, where

$$f(x) = \begin{cases} 1 & (x = 0), \\ x^{-1}\sin x + bx^2 & (0 \neq x \in I), \end{cases}$$

and

$$f'(x) = \begin{cases} 0 & (x = 0), \\ x^{-2}g(x) & (0 \neq x \in I), \end{cases}$$

where $g(x) = x\cos x - \sin x + 2bx^3$. Since the sine function is concave on I and $\sin 0 = 0$, $\sin(\pi/2) = 1 = 6b(\pi/2)$, it follows that $\sin x \geqslant 6bx$ $(x \in I)$, i.e., $g'(x) \leqslant 0$ $(x \in I)$. Hence g is decreasing and $g(0) = 0$ implies $g(x) \leqslant 0$ $(x \in I)$. Thus $f'(x) \leqslant 0$ $(x \in I)$ so that f is non-increasing and

$$a = f\left(\frac{\pi}{2}\right) = \frac{2}{\pi} + \frac{\pi}{12}.$$

Solution II *by Donald Caccia, DePaul University.* We have

$$\text{(1)} \qquad \sin\theta \geqslant \frac{2}{\pi}\theta + \frac{\theta}{\pi^3}(\pi^2 - 4\theta^2) \quad \text{in} \quad 0 \leqslant \theta \leqslant \frac{\pi}{2}.$$

Inequality (1) is slightly stronger than the proposed inequality and is sharp in the sense that $1/\pi^3$ cannot be replaced by a larger constant. Inequality (1) may be rewritten as

$$\text{(1')} \qquad \sin\theta \geqslant 3\frac{\theta}{\pi} - 4\frac{\theta^3}{\pi^3} \quad \text{in} \quad 0 \leqslant \theta \leqslant \frac{\pi}{2}.$$

Proof of (1'). Let

$$f(\theta) = \sin\theta - 3\frac{\theta}{\pi} + 4\frac{\theta^3}{\pi^3}, \quad I = \left[0, \frac{\pi}{4}\right], \quad J = \left[\frac{\pi}{4}, \frac{\pi}{2}\right].$$

Then $f^{(2)} \leqslant 0$ on $\partial I = \{0, \pi/4\}$, $f^{(4)} \geqslant 0$ on I implies $f^{(2)} \leqslant 0$ on I, whence f is concave on I so that $f \geqslant 0$ on ∂I implies $f \geqslant 0$ on I. On J, $f^{(1)}$ is convex since $f^{(3)}(\pi/4) > 0$ and $f^{(4)} > 0$ implies $f^{(3)} > 0$. Since $f^{(1)} \leqslant 0$ on ∂J and $f^{(1)}$ is convex on J, $f^{(1)} < 0$ on J. Hence $f(\pi/2) = 0$ implies $f \geqslant 0$ on J. ■

Proof of sharpness. Let $\varepsilon > 0$ be given. We must show

$$g(x) = g(x, \varepsilon) = f(x) - \varepsilon\theta(\pi^2 - 4\theta^2)$$

is not positive on $[0, \pi/2]$. But $g(\pi/2) = 0$ and $g'(\pi/2) > 0$, so g is negative slightly to the left of $\pi/2$. ■

Also solved by 59 other readers and the proposer.

The Existence of "Small" Prime Solutions of $x^{p-1} \not\equiv 1 \pmod{p^2}$

E 2956 [1982, 498]. *Proposed by Barry Powell, Kirkland, WA.*

Prove that if p is a prime, $p \equiv 1 \bmod 4$, there exists a prime $q < \sqrt{p}$ for which $q^{p-1} \not\equiv 1 \bmod p^2$. (See E 2435 [1973, 943].)

Solution by Nicholas Tzanakis, University of Crete, Iraklion, Crete, Greece. We shall prove the following more general result: Let $2 < m \equiv 1$ or $2 \pmod 4$ and suppose, moreover, that m has k distinct prime divisors congruent to 1 modulo 4 and no prime divisor congruent to 3 modulo 4. Then there exists a prime $q \leqslant \sqrt{m - 4^{k-1}}$ for which $q^{\phi(m)} \not\equiv 1 \pmod{m^2}$.

Proof. It is a well-known fact that there exist 2^{k-1} sets $\{x_i, y_i\}$, $1 \leqslant i \leqslant 2^{k-1}$, of positive integers x_i and y_i satisfying $m = x_i^2 + y_i^2$, $(x_i, y_i) = 1$ (see e.g. [1], Theorem 62, or [2], Chap. 3, §§10–12).

For any fixed i we must have either $x_i^{\phi(m)} \not\equiv 1$ or $y_i^{\phi(m)} \not\equiv 1 \pmod{m^2}$. Indeed, suppose the contrary. On putting $r = \phi(m)/2$ (r even), we have

$$y_i^{2r} = \left(x_i^2 - m\right)^r = x_i^{2r} - rx_i^{2(r-1)} \cdot m + m^2(\cdots),$$

so that $rx_i^{2(r-1)} \equiv 0 \pmod m$. Since $(x_i, m) = 1$ we must have $r \equiv 0 \pmod m$, clearly impossible.

Thus, without loss of generality, we may suppose that for every $i = 1, 2, \ldots, 2^{k-1}$ we have

$$x_i^{\phi(m)} \not\equiv 1 \pmod{m^2}. \tag{1}$$

We may also suppose that $y_1 > y_2 > y_3 > \cdots$, so that $y_1 \geqslant 2^{k-1}$.

On the other hand, in view of (1), x_1 must have a prime divisor q for which $q^{\phi(m)} \not\equiv 1 \pmod{m^2}$. The prime q also satisfies

$$q \leqslant x_1 = \sqrt{m - y_1^2} \leqslant \sqrt{m - 4^{k-1}}$$

and this completes the proof.

Note. It is not hard to see that we cannot have $[\sqrt{m - 4^{k-1}}] \leqslant [\sqrt{m-1}] - 2$. However, it may happen that $[\sqrt{m - 4^{k-1}}] = [\sqrt{m-1}] - 1$, as, e.g., when $m = 32045 = 5 \cdot 13 \cdot 17 \cdot 29$.

References

1. L. E. Dickson, Introduction to the Theory of Numbers, Dover, New York, 1957.
2. Hansraj Gupta, Selected Topics in Number Theory, Abacus Press, Kent, 1980.

Also solved by J. Ampe, K. L. Bernstein, R. Breusch, K. Lau, L. Yuen and L. Pei (Hong Kong), O. P. Lossers (The Netherlands), L. E. Mattics, K. Rogers, F. W. Schmidt and R. Simion, R. E. Shafer, University of South Alabama Problem Group, and G. H. Wenzel (West Germany).

Arcs with Positive Two-dimensional Lebesgue Measure

E 2975 [1982, 756]. *Proposed by F. Burton Jones, University of California, Berkeley.*

An arc is a one-to-one reversibly continuous image of [0,1] of the real numbers. Does there exist an arc in the plane whose 2-dimensional Lebesgue measure is positive?

The answer is yes. Several readers provided references to this result, the most widely cited reference being *Counterexamples in Analysis* by Gelbaum and Olmsted, pages 135–138.

An Infinite Product Involving the Euler Totient Function

E 2985 [1983, 55]. *Proposed by P. J. Forrester, University of Melbourne, and M. L. Glasser, Clarkson College.*

Show that for $0 < x < 1$

$$\prod_{\substack{k=1\\ \text{odd}}}^{\infty} [(1+x^k)/(1-x^k)]^{\phi(k)/k} = \exp[2/(x^{-1}-x)]$$

where $\phi(k)$ is Euler's totient.

Solution by H. L. Abbott, University of Alberta, Canada. Let

$$F(x) = \prod_{\substack{k=1\\ k \text{ odd}}}^{\infty} \left(\frac{1+x^k}{1-x^k}\right)^{\phi(k)/k}.$$

Then

$$\begin{aligned}\log F(x) &= \sum_{\substack{k=1\\ k \text{ odd}}}^{\infty} \frac{\phi(k)}{k} \log\left(\frac{1+x^k}{1-x^k}\right)\\ &= 2\sum_{\substack{k=1\\ k \text{ odd}}}^{\infty} \frac{\phi(k)}{k} \sum_{m=0}^{\infty} \frac{x^{k(2m+1)}}{2m+1}\\ &= 2\sum_{k=0}^{\infty}\left(\sum_{d|2k+1}' \phi(d)\right)\frac{x^{2k+1}}{2k+1}\\ &= 2\sum_{k=0}^{\infty} x^{2k+1}\\ &= \frac{2x}{1-x^2},\end{aligned}$$

where we have used the identity $\sum_{d|n} \phi(d) = n$. Thus $F(x) = \exp(2x/(1-x^2))$.

All of the manipulations are easily justified since the product and series converge absolutely for $-1 < x < 1$.

Also solved by 26 other readers and the proposers.

V. Mascioni (Switzerland) has noted that the identities

$$\prod_{k=1}^{\infty} (1-x^k)^{\phi(k)/k} = \exp[-x/(1-x)], \quad \prod_{k=1}^{\infty} (1-x^k)^{\mu(k)/k} = e^{-x},$$

are in Pólya and Szegő, *Problems and Theorems in Analysis* II, (Chap. VIII) 72.1. The latter appeared as problem 4072 in this MONTHLY, 50 (1943) pp. 124–125, proposed by Richard Bellman. The first identity immediately solves the problem on replacing x by $-x$ and combining. A. Mercier (Canada) has obtained the generalization, for $-1 < x < 1$,

$$\prod_{\substack{k=1\\ \text{odd}}}^{\infty} \left(\frac{1+x^k}{1-x^k}\right)^{f(k)/k} = \exp\left[2\prod_{\substack{k=1\\ \text{odd}}}^{\infty}\left(\sum_{d|k} f(d)\right)\frac{x^k}{k}\right].$$

R. Sivaramakrishnan (India) has given a further generalization for products $\prod(1+x^k)^{g(k)/k}(1-x^k)^{f(k)/k}$, and has pointed out that the case when $g(k)=0$ has appeared before in N. Balasubramanian, Mathematics Student, 19 (1961) pp. 89–92.

A Limit Obtained from an Increasing Sequence

E 3000 [1983, 335]. *Proposed by P. Erdős, Hungarian Academy of Sciences.*

Let $0 < u_1 < u_2 < \cdots$ be an infinite sequence; suppose $\Sigma 1/u_i$ converges. Denote by $f(x)$ the number of pairs (i, j) for which the partial sum $\Sigma_i^j u_r$ is $\leqslant x$. Prove that $\lim f(x)/x = 0$ $(x \to \infty)$.

Solution I *by J. B. Wilker, University of Toronto, Canada.* Let $f(x) = \Sigma_1^\infty f_i(x)$ where $f_i(x)$ denotes the number of pairs (i, j) with $i \leqslant j$ and $\Sigma_i^j u_r \leqslant x$.

Since u_r is increasing, $f_i(x) \leqslant x/u_i$ and the equation $\Sigma_1^\infty 1/u_i = S$ yield $f(x)/x \leqslant S$. To show that $\lim_{x\to\infty} f(x)/x = 0$ we use further properties of the functions $f_i(x)$ together with the well-known result on monotone series that $\lim_{n\to\infty} n \cdot 1/u_n = 0$.

Since there are $(j - i)$ pairs (i, k) with $k < j$, $f_i(x) \leqslant (j - i) + f_j(x - \Sigma_i^{j-1} u_r)$ and since f_i is monotone increasing in x, $f_i(x) \leqslant (j - i) + f_j(x)$. For any value of n we can write

$$\begin{aligned} f(x) &= \sum_1^{n-1} f_i(x) + \sum_n^\infty f_i(x) \\ &\leqslant \sum_1^{n-1} [(n - i) + f_n(x)] + \sum_n^\infty f_i(x) \\ &= \frac{n(n-1)}{2} + nf_n(x) + \sum_{n+1}^\infty f_i(x) \\ &\leqslant \frac{n(n-1)}{2} + nx/u_n + \sum_{n+1}^\infty x/u_i. \end{aligned}$$

Thus

$$f(x)/x \leqslant \frac{n(n-1)}{2x} + n \cdot 1/u_n + \sum_{n+1}^\infty 1/u_i.$$

The last two terms can be made arbitrarily small by taking n large enough and then the first term can be made arbitrarily small by taking x large enough.

Solution II *by L. E. Mattics, University of South Alabama.* For $x \geqslant u_1$ let $g(x)$ be the largest integer such that $u_{g(x)} \leqslant x$. Since $0 < u_1 < u_2 < \cdots$ and $\Sigma 1/u_i$ converges, it is easy to show that $\lim g(x)/x = 0$ $(x \to \infty)$. For positive integers n with $n \leqslant g(x)$, let $f(x, n)$ be the largest integer such that

$$\sum_{r=n}^{f(x,n)+n-1} u_r \leqslant x;$$

then

$$f(x) = \sum_{n=1}^{g(x)} f(x, n).$$

Given $\varepsilon > 0$, there is a positive integer N such that $\sum_{r=N}^\infty 1/u_r < \varepsilon$. For $1 \leqslant n \leqslant g(x)$

$$\begin{aligned} f(x, n) &= \sum_{r=n+N-1}^{f(x,n)+n-1} u_r(1/u_r) + N \\ &\leqslant N + x(1/u_{N+n-1}), \end{aligned}$$

so

$$f(x) \leqslant Ng(x) + x \sum_{r=N}^{g(x)+N} (1/u_r) < Ng(x) + \varepsilon x.$$

Hence, $f(x)/x = o(1)$ as $x \to \infty$.

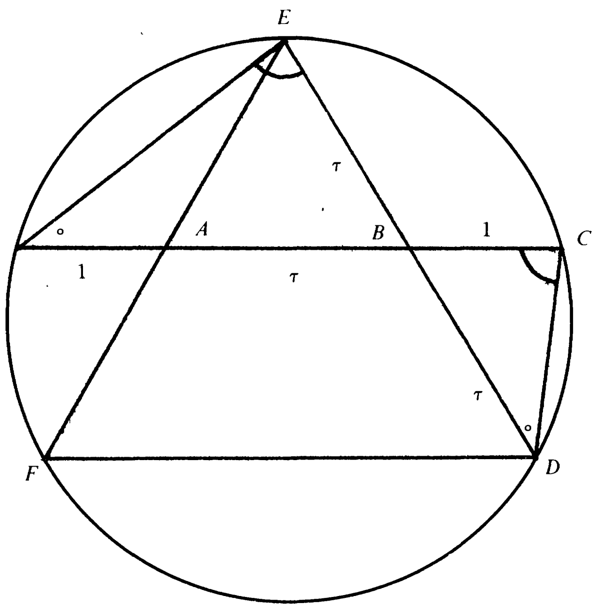
Also solved by R. Breusch, R. O. Davies (England), R. Goldstein, A. Kovaćec (Austria), and C. A. Meyer (Switzerland).

The Golden Ratio from an Equilateral Triangle and Its Circumcircle

E 3007 [1983, 482]. *Proposed by George Odom, Poughkeepsie, NY.*

Let A and B be the midpoints of the sides EF and ED of an equilateral triangle DEF. Extend AB to meet the circumcircle (of DEF) at C. Show that B divides AC according to the golden section.

Solution by Jan van de Craats, The Netherlands.



$\tau^2 = \tau + 1$

Also solved by 74 other readers and the proposer.

Separated Subsets of a Connected Space

E 3011 [1983, 483]. *Proposed by C. P. Popescu, Romania.*

Let X be a connected (path connected) topological space and let A, B be two nonvoid disjoined subsets of X such that boundary (A) = boundary (B). Is it true that $\bar{A} \cup \bar{B} = X$?

Solution by Jay Hook, Florida International University, Miami, FL. No. Let X be the union of the x- and y-axes in the plane, let A be the positive x-axis, and let B be the positive y-axis.

Also solved by forty other readers and the proposer.

An Inequality Involving Powers

E 3016 [1983, 567]. *Proposed by Eugene Levine, Adelphi University.*

Let $a_1, a_2, \ldots, a_n$ be positive numbers ($n > 2$), let $s = a_1 + a_2 + \cdots + a_n$, and let $0 < \beta \leqslant 1$. Prove that

$$\sum_{k=1}^{n}\left(\frac{s-a_k}{a_k}\right)^{\beta} \geqslant (n-1)^{2\beta}\sum_{k=1}^{n}\left(\frac{a_k}{s-a_k}\right)^{\beta}.$$

Further, show that equality holds if and only if $a_1 = a_2 = \cdots = a_n$.

Solution by Aage Bondesen, Denmark. Define

$$A = \sum_{k=1}^{n}\left(\frac{s-a_k}{a_k}\right)^{\beta} = (n-1)^{\beta}\cdot\sum_{k=1}^{n}\frac{1}{a_k^{\beta}}\cdot\left(\frac{1}{n-1}\sum_{\substack{i=1\\ i\neq k}}^{n}a_i\right)^{\beta}.$$

Since the function $x \to -x^{\beta}$ is convex for x positive, we have for $k = 1, 2, \ldots, n$ (Jensen's inequality)

$$-\left(\frac{1}{n-1}\sum_{\substack{i=1\\ i\neq k}}^{n}a_i\right)^{\beta} \leqslant -\frac{1}{n-1}\sum_{\substack{i=1\\ i\neq k}}^{n}a_i^{\beta}, \quad \text{or equivalently} \quad \left(\frac{1}{n-1}\sum_{\cdots}a_i\right)^{\beta} \geqslant \frac{1}{n-1}\sum_{\cdots}a_i^{\beta},$$

with equality if and only if $a_1 = \cdots = a_{k-1} = a_{k+1} = \cdots = a_n$. Therefore

$$A \geqslant (n-1)^{\beta}\cdot\sum_{\substack{i,k=1\\ i\neq k}}^{n}\frac{1}{n-1}a_i^{\beta}a_k^{-\beta}$$

with equality if and only if $a_1 = a_2 = \cdots = a_n$. We transform this into

$$A \geqslant (n-1)^{\beta}\cdot\sum_{i=1}^{n}a_i^{\beta}\cdot\frac{1}{n-1}\sum_{\substack{k=1\\ k\neq i}}^{n}a_k^{-\beta}.$$

Now the function $x \to x^{-\beta}$ is convex for x positive, so (Jensen again)

$$A \geqslant (n-1)^{\beta}\cdot\sum_{i=1}^{n}a_i^{\beta}\cdot\left(\frac{1}{n-1}\sum_{\substack{k=1\\ k\neq 1}}^{n}a_k\right)^{-\beta} = (n-1)^{\beta}\cdot\sum_{i=1}^{n}\frac{(n-1)^{\beta}\cdot a_i^{\beta}}{(s-a_i)^{\beta}},$$

with equality if and only if $a_1 = \cdots = a_n$. We conclude that

$$A \geqslant (n-1)^{2\beta}\cdot\sum_{i=1}^{n}\left(\frac{a_i}{s-a_i}\right)^{\beta},$$

with equality if and only if $a_1 = \cdots = a_n$, as was to be proved.

Also solved by University of South Alabama Problem Group and the proposer.

ADVANCED PROBLEMS

For instructions about submitting solutions of these Advanced Problems, which should be mailed by January 31, 1987, see the inside front cover. The solver's full post-office address should be on each sheet.

6524. *Proposed by Gérard Letac and Guy Yétérian, Université Paul Sabatier, Toulouse, France.*

Let p and q be positive numbers, and $\{X_n\}_{n=0}^{\infty}$ a sequence of independent random variables

with the same distribution

$$\beta_{p,q}(dx) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1}(1-x)^{q-1}\,dx$$

on [0, 1]. Find the distribution of

$$\sum_{n=0}^{\infty} (-1)^n X_0 X_1 \cdots X_n.$$

6525. *Proposed by Andrew Lenard, Indiana University.*

Let S be an open subset of the complex plane. Prove that a function f is holomorphic on S if and only if

$$\sup_D{}^* \frac{\left|\sum_{j=1}^{n} (z_{j+1} - z_j) f(z_j)\right|}{\sum_{j=1}^{n} |z_{j+1} - z_j|^2} < \infty$$

for every closed disc $D \subset S$. Here the star indicates that the supremum is taken over all finite sequences $z_1, z_2, \ldots, z_n, z_{n+1}$ of points of D (n arbitrary) such that $z_{n+1} = z_1$ and at least two of the z_j are distinct.

SOLUTIONS OF ADVANCED PROBLEMS

Resource Sharing for Efficiency

6482* [1984, 652; 1986, 488] (**Addendum**).

If $B(p, a) = [aR(a, p)]^{-1}$, where $R(a, p)$ is the integral specified in the problem, then $B(p, a)$ is the Erlang blocking formula in queueing or traffic theory. This problem involved an inequality concerning $R(a, p)$ and $R(a, q)$. Referring to the printed solution, we find that the case of integer p and q is established by different methods in D. R. Smith and W. Whitt, *Resource sharing for efficiency in traffic systems*, Bell System Tech. J., 60 (1981), 39–55; also see E. Arthurs and B. W. Stuck, *Problem* 80-19*, SIAM Review, 23 (1981), 527–528. The solution for general p and q follows from the convexity of $B(p, a)$ in p using an argument by H. Shulman reproduced on p. 54 of Smith and Whitt. This convexity had been widely conjectured, but not established. However, while this issue was being prepared for publication, the solvers notified the editor that the convexity was established by A. A. Jagers and E. A. van Doorn, *On the continued Erlang loss function*, Operations Research Letters, to appear. Hence, problem 6482 is established for arbitrary positive p and q.

A Metric for the Symmetric Group

6486 [1985, 62]. *Proposed by Liviu I. Nicolaescu (student), University Ai Cuza Iasi, Romania.*

Let S_n denote the set of all permutations of $1, 2, \ldots, n$. For $\sigma, \tau \in S_n$, define $d(\sigma, \tau)$ to be the number of inversions of the permutation $\sigma \cdot \tau^{-1}$. Prove that d is a metric on S_n.

Solution by Michael Josephy, Universidad de Costa Rica, San José, Costa Rica, and Jean-Charles Leccia, Aix-en-Provence, France (independently). Let $\|\sigma\|$ be the number of inversions of $\sigma \in S_n$. This is the cardinality of $I(\sigma)$, where

$$I(\sigma) = \{(i, j):\ 1 \le i < j \le n,\quad \sigma i > \sigma j\}.$$

To show that $d(\sigma, \tau) = \|\sigma\tau^{-1}\|$ is a metric, it suffices to prove for all $\sigma, \tau \in S_n$ that

(1) $\|\sigma\| \geq 0$ and $\|\sigma\| = 0$ if and only if $\sigma = 1$,
(2) $\|\sigma\| = \|\sigma^{-1}\|$,
(3) $\|\sigma\tau\| \leq \|\sigma\| + \|\tau\|$.

The first two are easy. For the third write

$$I(\sigma\tau) = \{(i, j): i < j, \quad \tau i < \tau j, \quad \sigma\tau i > \sigma\tau j\}$$
$$\cup \{(i, j): i < j, \quad \tau i > \tau j, \quad \sigma\tau i > \sigma\tau j\}.$$

The first set in the union can be embedded in $I(\sigma)$, while the second is a subset of $I(\tau)$.

Leccia notes that the following exercise is in Luc Moisotte, 1850 *Exercices de Mathèmatiques* (Dunod-Université). Let $N(\sigma)$ be the set

$$\{i : \sigma(i) \neq i\}.$$

Then $|N(\sigma\tau^{-1})|$ is a metric on S_n (generally different, of course, from $d(\sigma, \tau)$).

Gerhard Behrendt (Federal Republic of Germany), William P. Wardlaw (U.S. Naval Academy), and R.H. Jeurissen (The Netherlands) connected Nicolaescu's problem with translation invariant metrics, group norms, and certain bipartite graphs, respectively.

Also solved by Anders Bager (Denmark), Armin Barth (Switzerland), Gerhard Behrendt (Federal Republic of Germany), Steve Benzel, F. S. Cater, Steven Davis, Crist Dixon, S. H. E. Hou, A. A. Jagers (The Netherlands), R. H. Jeurissen (The Netherlands), O. P. Lossers (The Netherlands), Ondrej Mateus (Czechoslovakia), Jean-Marie Monier (France), Victor Pambuccian (Romania), Dennis Spellman, Carol Stanton, John Henry Steelman, William P. Wardlaw, Pei Yuan Wu (Taiwan), and the proposer.

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER

Differential Geometry and Relativity Theory. By Richard L. Faber. Marcel Dekker, 1983. x + 255 pp.

Semi-Riemannian Geometry. By Barrett O'Neill. Academic Press, 1983. xiii + 468 pp.

JOHN A. THORPE
*Department of Mathematics, SUNY at Stony Brook, Stony Brook, NY 11794**

Einstein's theory of relativity has a special fascination for nearly everyone. It captures the imagination of the scientific layperson because of its reputation for complexity and its apparent paradoxes. It is the stuff that good science fiction is made of. Its implications lead to philosophical and even to religious discussions on the nature of the universe and of life itself. For a physicist, of course, it is one of the cornerstones of modern scientific theory. For a geometer like myself, it is an exciting example of the applicability of geometry to physical science.

Indeed, general relativity *is* a geometric theory. It postulates a geometric model of the universe (space and time) in which the effects of gravity are built into the geometry in such a way that objects moving freely in the gravitational field will travel along the straight lines, or geodesics, of the geometry. The geometry is not the Euclidean geometry that we learn in secondary school. Nor

*Currently on leave at the National Science Foundation, Washington, DC 20550.

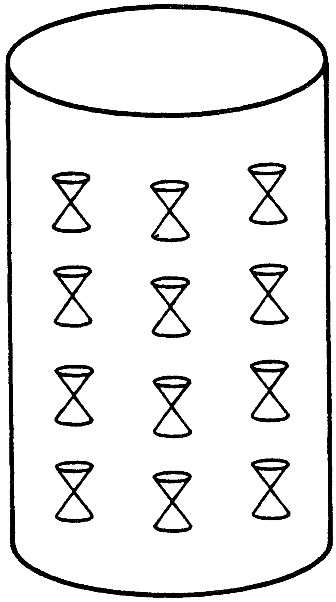


FIG. 2

Figs. 2 and 3 portray two cylindrical models of spacetime (two dimensions suppressed). The first of these (Fig. 2) is a quite reasonable model of a spacetime (a closed space model). In the figure this spacetime appears as the Cartesian product of a circle and a line. In four dimensions it would be the Cartesian product of a 3-sphere and a line. The second model (Fig. 3), which also appears in the figure to be the product of a line and a circle, would in four dimensions be the product of Euclidean 3-space and a circle. This model is usually regarded as unreasonable because of the presence of closed worldlines. If, for example, the circle in Fig. 3 represented the worldline of an individual, then that individual would eventually return to the same point in spacetime that he had occupied before (same "point in space" at the same "time" that he was previously there). In particular, he would enter his own past and conceivably would be able to change history!

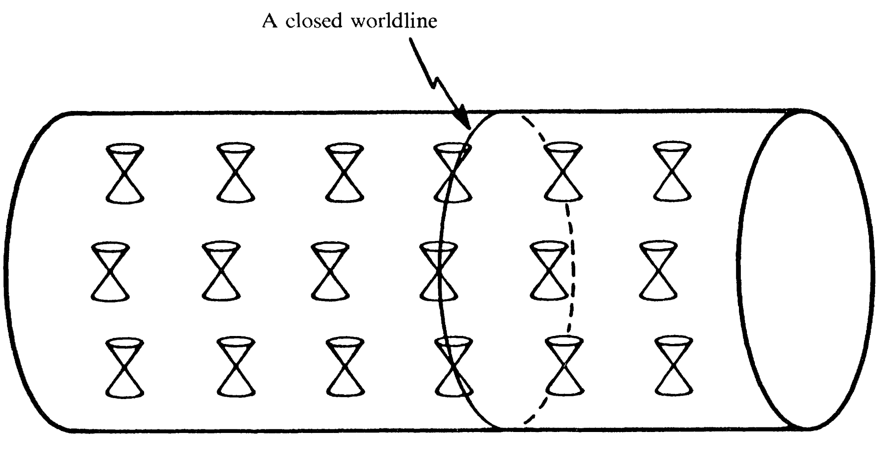


FIG. 3

Some potentially nice models for spacetime are not only unreasonable but are actually impossible. The four dimensional sphere, for example, cannot be given a causal structure. (Nor can the 2-sphere, although the 3-sphere can!) The reason is the same as the reason that there is no

continuous nowhere zero vector field on the 2-sphere, or on the 4-sphere (the famous theorem, "you can't comb the hair on a coconut").

A causal structure (field of light cones) on a manifold nearly determines a Lorentz metric on the manifold. The only missing information is a scale factor (the causal structure determines the Lorentz metric up to conformal equivalence). For much of relativity theory, only the causal structure is needed, and conceptually it is much simpler. But this brief discussion would be incomplete without a definition of Lorentz metric.

A *Lorentz metric* on a manifold is a field of nonsingular inner products with the Lorentz signature $(- - - +)$. (At each point of the manifold there is a vector space, called the tangent space. A Lorentz metric is an assignment of a symmetric bilinear form, with signature $(- - - +)$, to each of these tangent spaces.) Given a Lorentz metric, with inner product $\langle , \rangle$ at a point p, the light cone at p consists of those vectors v with $\langle v, v \rangle = 0$.

The two books under review are attempts to bring this fascinating theory within the grasp of upper division undergraduate and first year graduate students of mathematics.

The book by Faber, written for undergraduates, devotes its first third to a rather classical and somewhat sketchy treatment of the differential geometry of surfaces, its middle third to a discussion of special relativity, and its final third to an introduction to general relativity. Included are discussions of perihelion precession and of the bending of light. This book should be especially useful for students who are taking a standard undergraduate course in differential geometry and wish to gain some insight into how the subject interfaces with physics and astronomy.

The book by O'Neill is written for a first graduate course in differential geometry. This book is more ambitious as well as more advanced than the one by Faber. It treats all of the standard topics that are normally covered in a first graduate course in differential geometry but it includes from the outset both the Riemannian and the Lorentzian cases. The treatment of general relativity is confined to the final three of the fourteen chapters, although the terminology of the subject is used freely throughout the text and one chapter in the middle of the text is devoted to special relativity. Among the topics in relativity that appear in O'Neill's book are cosmological models and the singularity theorems of Hawking and Penrose. There is a detailed chapter on Schwarzschild geometry, which includes discussions of perihelion precession, bending of light, and black holes. This book is beautifully written and deserves a place on the bookshelf of every practicing, and every aspiring, differential geometer.

Lorentzian geometry is the geometry of the world in which we live. In the small (that is, as long as distances and velocities are small), Euclidean geometry and Newtonian physics can suffice. But if we look outward into the universe, and consider large scale phenomena, then we must come to grips with Lorentzian geometry. These two books will help us to do that.

Computer Mathematics. By D. J. Cooke and H. E. Bez. Cambridge University Press, 1984.

ALBERT L. RABENSTEIN
Department of Mathematics, Washington and Jefferson College, Washington, PA 15301

What should be the role of discrete mathematics in the undergraduate mathematics curriculum? Primarily because of the increased interest in computer science, this question has received considerable attention in recent years, including articles in the MONTHLY and *The College Mathematics Journal* and discussions and lectures at MAA meetings.

Most of the discussion has centered on courses at the freshman level, either replacing or supplementing the standard calculus sequence. (See, for example, the November 1984, issue of *The College Mathematics Journal*.) There are, however, other possibilities for a course in discrete

mathematics, and the book under review is aimed at a more advanced audience of undergraduates which the authors describe as "honours students in the Department of Computer Studies..." at universities in Great Britain.

At Washington and Jefferson College we have also considered a more advanced course, and have decided to offer a junior-level course in discrete structures, which will replace a standard course in abstract algebra. The main reasons for the change were the relatively small number of traditional mathematics majors and the creation of a new computer science option within the mathematics department. The new course will be required for both the "pure" mathematics majors and those pursuing the computer science option, largely because the number of mathematics majors is insufficient to justify separate courses. The new course, together with a course in linear algebra, will constitute the "nonanalysis" part of the mathematics major; regular majors will still be required to take advanced calculus.

Even if we had not created the new computer science track, the course change would still be appealing. Since our ordinary mathematics majors do not become research mathematicians, consideration of their long term career aims makes a course that is more applications oriented than the usual abstract algebra course more desirable. On the other hand, some majors do seek teacher certification, and hence they need an acquaintance with algebraic structures such as groups, rings, and fields. Thus it was necessary to include some of this standard material in our new course.

The design of the course in discrete structures was based on the recommendations of the CUPM report "Recommendations for a General Mathematical Sciences Program," published by the MAA in 1981. Its content will include groups and other algebraic structures, combinatorics, and graph theory. As indicated above, the abstract algebra portion is necessary for prospective secondary school teachers. The combinatorics and graph theory segments are important for computer science, but are also applicable in other areas as well, as indicated in the CUPM report. We also include fundamental material on sets, relations, and logic. Other topics will be selected from boolean algebra, finite state machines, and formal languages. These certainly have applications in computing, but no great knowledge of computing is required for an understanding, and they can be rewarding to regular mathematics majors.

There are, of course, some disadvantages to postponing a discrete mathematics course until the junior year. In particular, a few specialized topics that are usually proposed for a more elementary course, such as postfix notation, will have to be incorporated in regular computer science courses.

Computer Mathematics is one of a number of textbooks recently published that cover the topics mentioned above with varying degrees of emphasis. Because parts of it assume a fair amount of mathematical sophistication on the part of the reader, the book would not be suitable for beginning students in this country. This book covers the topics discussed above, although logic and combinatorics are not developed in depth. It also includes chapters on finite arithmetic and computer geometry, considerable material on linear algebra, and a review of single variable calculus. The scope is large, the presentation often terse. The authors tend to introduce new topics with definitions that are general and abstract. Precision is enhanced, but introductory motivation is often omitted. The emphasis throughout the book is on the mathematics rather than the applications, although a few applications are included. The mathematics is of interest in its own right, in part because it deals with such fundamental ideas.

The reader who masters the material in this book will certainly be well-prepared mathematically to study computer science. The authors develop, in a logical, interesting way, a large (probably too large) body of theory. But the instructor who teaches from this text will have to provide extra insight and motivation.

The level and content of courses in discrete mathematics have not yet been standardized. The book under review and the course described above suggest an alternative to the freshman level course most often proposed.

Introduction to Topology. By T. W. Gamelin and R. E. Greene. Saunders College Publishing, New York, 1983, xii + 196 pp.

Basic Topology. By M.A. Armstrong. Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1983, xii + 251 pp.

ISAAC NAMIOKA
Department of Mathematics, University of Washington, Seattle, WA 98195

"What do you teach?" is the question I am often asked at parties. My answer, "I teach mathematics," invariably elicits the next question, "What kind of mathematics?"

When I answer, "Calculus and general topology," for some reason people are not interested in calculus (we ought to do something about that), but they instinctively bear down on topology: What is T-O-P-O-L-O-G-Y?

The official answer to this question goes something like this: topology is a study of those geometric properties that are invariant under continuous changes like stretching and bending. This sounds too much like a dictionary definition. I prefer the Old Testament ring of the following passage from Lefschetz [**5**] (which we might call the wisdom of Solomon): "Topology begins where sets are implemented with some cohesive properties enabling one to define continuity. The sets are called topological spaces."

However, such answers are not enlightening to my inquirers. I usually extricate myself by mentioning the seven bridges in the city of Königsberg (now Kaliningrad, U.S.S.R.; see p. 663 of the December, 1984, issue of this MONTHLY), an ant crawling on a Möbius band, the Euler characteristic, and the definition of a topologist as a man who can't distinguish a doughnut from a coffee cup (see p. 88 of Kelley [**4**]). But of course I am doing injustice to topology. It is much more than that.

Topology as a discipline goes back to Riemann, Poincaré, and Cantor, all nineteenth century mathematicians. Thus, it is a relatively young field, but it has been so intensively developed that its influence now permeates every branch of mathematics. The language of topology has become indispensable in such diverse areas as number theory, ordinary and partial differential equations, operator theory, and abstract algebra. Indeed, the influence of topology is seen even in areas outside mathematics, in such fields as solid state physics, cosmology, particle physics, and quantum chemistry.

Traditionally we distinguish two divisions within topology: general (or point set) topology and algebraic topology. They have usually been taught separately using different textbooks. Thus general topology is often slanted toward analysis. J.L. Kelley in his popular *General Topology* [**4**] states in the preface that "I have, with difficulty, been prevented by my friends from labeling it [the book]: What Every Young Analyst Should Know." For instance, the Brouwer fixed point theorem is not even mentioned in Kelley's book. I suppose he considers it to be a theorem in algebraic topology. A sort of demarcation of labor prevails here. The development of mathematics in the last decade or two indicates, however, that even analysts cannot be indifferent to algebraic topology. As an illustration, I should like to describe how the Brouwer fixed point theorem entered the invariant subspace problem.

Let n be an arbitrary positive integer. The Brouwer fixed point theorem states that any continuous function f from the unit ball $D_n = \{x \in \mathbb{R}^n\colon \sum_{i=1}^n x_i^2 \leqslant 1\}$ in euclidean n-space to itself has a fixed point, i.e., a point x_0 in D_n such that $f(x_0) = x_0$. Schauder generalized it to the infinite dimensional case: If C is a compact convex subset of a Banach space, then each continuous function from C to C has a fixed point. The proof is quite easy once the finite dimensional case (i.e., the Brouwer theorem) is granted.

Now the invariant subspace problem. It has long been an outstanding problem to decide which continuous linear operators $T\colon E \to E$ (where E is either a Hilbert space or a complex Banach space) admit a proper invariant subspace, i.e., a closed linear subspace F of E such that $\{0\} \neq F \neq E$ and $T(F) \subset F$. In the 1930s, von Neumann proved that each compact operator on

a Hilbert space admits a proper invariant subspace, but the proof was never published. (An operator T: $E \to E$ is said to be compact if the image under T of the unit ball has a compact closure.) In a paper published in 1954 [**1**], N. Aronszajn and K. T. Smith generalized the result of von Neumann to Banach spaces. Smith then asked if an operator T on a Banach space such that T^2 is compact would admit a proper invariant subspace. This question was answered in the affirmative for Hilbert spaces by A.R. Bernstein and A. Robinson in [3] (1966) and for Banach spaces by Bernstein in [**2**] (1967). They actually established a stronger result: If T is an operator on a Hilbert space (or complex Banach space) such that $p(T)$ is compact for some polynomial p, then T admits a proper invariant subspace. Their papers are not short, and one has to know a lot about non-standard analysis in order to understand them. Then came a young Russian mathematician, V. I. Lomonosov, who in 1973 gave a breathtakingly elegant proof of a theorem far stronger than any of the above [**6**]. Lomonosov proved that if T is an operator on an infinite dimensional Banach space E such that $TK = KT$ for some non-zero compact operator K on E, then T admits a proper invariant subspace. His proof, which is less than one page long complete with details, relies on the Schauder fixed point theorem! I can still recall vividly the occasion when a friend from Berkeley pulled from his pocket a crumpled sheet of paper with Lomonosov's proof and said that it was making a lot of people in Berkeley "unhappy."

The two books under review are both introductory books on topology in the broader sense, i.e., each has components of both point set and algebraic topology. Both of them are accessible to a student who has had undergraduate real analysis and some abstract algebra. However, the styles of the two authors (actually three authors, because one is written jointly) are quite different. By far the livelier of the two is *Basic Topology* by M. A. Armstrong. Since the author describes his idea for the book very well, I will let him speak. "The layout is as follows. There are ten chapters, the first of which is a short essay intended as motivation. Each of the other chapters is devoted to a single important topic, so that identification spaces, simplicial homology, knots, and covering spaces, all have a chapter to themselves.

"Some motivation is surely necessary. A topology book at this level which begins with a set of axioms for topological spaces, as if these were an integral part of nature, is in my opinion doomed to failure. On the other hand, topology should not be presented as a collection of party tricks (colouring knots and maps, joining houses to public utilities or watching a fly escape from a Klein bottle). These things all have their places, but they must be shown to fit into a unified mathematical theory, and not remain dead ends in themselves. For this reason, knots appear at the end of the book, and not at the beginning. It is not the knots which are so interesting, but rather the variety of techniques needed to deal with them."

The topics covered in *Basic Topology* include: The Tietze extension theorem (for metric spaces), the Jordan curve theorem in its weak form (the complement of a Jordan curve in the plane is not connected), the classification of combinatorial surfaces, invariance of simplicial homology groups under homeomorphisms, the Euler characteristic, the Lefschetz fixed point theorem (a generalization of the Brouwer fixed point theorem). After each section there is a good collection of problems, and the book is full of appealing drawings. With its conversational style, Armstrong's *Basic Topology* is a very enjoyable introduction to topology. But, one does not get much idea of how topology is used outside of topology proper. For that we must turn to the other book.

In *Introduction to Topology*, the authors T. W. Gamelin and R. E. Greene are very much concerned with the interaction of topology and other fields. The first half of this relatively slim book is devoted to general topology, and the second half to algebraic topology. The serious tone of the book makes the Armstrong book look like fun-and-games. *Introduction to Topology* is an introduction in the sense that the number of topics is limited. However, each topic the authors decide to include is treated fully. Thus Urysohn's lemma and the Tietze extension theorem are proved for normal spaces, and the Tychonoff theorem is proved in full generality via Alexander's lemma.

The book begins with a discussion of the topology of metric spaces. Here one finds how the abstract theorems in metric spaces are used to prove such theorems as the principle of uniform boundedness, the Cauchy-Picard theorem on the existence of solutions of differential equations, and the implicit function theorem for Banach spaces. This is followed by a chapter on general topological spaces where the name-brand theorems of the above paragraph are proved.

In the second half of *Introduction to Topology*, the authors choose homotopy theory as an illustration of algebraic methods. The actual amount of homotopy theory is meager, but there again the emphasis is on the interaction of topology with other areas. Thus there is a topological proof of the fundamental theorem of algebra, and the non-contractibility of the n-sphere (which is equivalent to the Brouwer fixed point theorem) is proved with the help of integration. The Jordan curve theorem is proved by a method similar to the one in *Basic Topology*, but the extra argument here proves the stronger conclusion: the complement of a Jordan curve in the plane has two components. At the end of each section, a hefty set of problems is provided.

The amount of material in *Introduction to Topology* is too much to cover in one term (as the authors themselves admit), and too little in one year. The authors sensibly suggest possible one-term introductory topology courses (for advanced undergraduate students and beginning graduate students) by selecting various parts of the book. On the other hand, *Basic Topology* would be a wonderful textbook for a year-long topology course for advanced undergraduate students.

Unfortunately topology as such is not usually taught at the undergraduate level, and the mathematics students I see around here are all busy satisfying distribution requirements and taking computer science courses. Because of the growing influence of topology on many areas of mathematics, however, it should be in the basic curriculum, at least for mathematics majors.

References

1. N. Aronszajn and K. T. Smith, Invariant subspaces of completely continuous operators, Ann. of Math., (2) 60 (1954) 345–350.

2. A. R. Bernstein, Invariant subspaces of polynomially compact operators on Banach spaces, Pacific J. Math., 21 (1967) 445–464.

3. A. R. Bernstein and A. Robinson, Solution of an invariant subspace problem of K. T. Smith and P. R. Halmos, Pacific J. Math., 16 (1966) 421–431.

4. J. L. Kelley, General Topology, Graduate Texts in Mathematics 27, Springer-Verlag, New York, Heidelberg, Berlin, 1975.

5. S. Lefschetz, Introduction to Topology, Princeton Mathematical Series No. 11, Princeton University Press, 1949.

6. V. I. Lomonosov, Invariant subspaces of the family of operators that can commute with a completely continuous operator (Russian), Funkcional Anal. i Priložen, 7 (1973) No. 3, 55–56.

LETTERS TO THE EDITOR

For instructions about submitting letters for publication in this department see the inside front cover.

Editor:

I think that Harley Flanders' review of Loren Larson's *Problem Solving through Problems* in the November 1985 issue should not go unanswered.

Though it would be possible to be as picky with the review as Professor Flanders was with the book, I will restrict myself to what I see as his main point, namely that Professor Larson's solutions could often be "better" and are sometimes "long, cloddish, and computational."

Contrary to Professor Flanders, I think this is a point in the book's favor. The book is

intended for students, not professionals, and methods of solution which seem natural to Professor Flanders would seem to be unnatural tricks to novice problem solvers and could easily cause them to think, "I could never have thought of that, it is hopeless, I am a failure, I will never be able to solve problems." On the other hand, some novice problem solvers *can* carry out computational solutions, and the glow of satisfaction gained more than outweighs the elegance lost. When I was an undergraduate I solved a MONTHLY problem by enumerating something like 24 cases: it was not the best solution, it was not published, and I would not do that now, but one must start somewhere.

Of course Professor Larson's book has flaws, as what book does not, but I think Professor Flanders goes too far in implying that it is a *bad* book. It may not be perfect, but it is valuable.

Underwood Dudley
Mathematics Department
DePauw University
Greencastle, IN 46135

---

174. MISCELLANEA

Handicapping Bodyweights

JOSEPH A. GALLIAN
Department of Mathematical Sciences, University of Minnesota, Duluth, MN 55812-2496

MONTHLY readers may be amused by the curious formula* below which I recently discovered while browsing through the *Handbook of Powerlifting*.

BODYWEIGHTS BETWEEN 40–126 KGS.

$$\begin{aligned} SF = {} & 0.631926\,exp(+01) \\ & -0.262349\,exp(+00)\,(BW) \\ & +0.511550\,exp(-02)\,(BW)^2 \\ & -0.519738\,exp(-04)\,(BW)^3 \\ & +0.267626\,exp(-06)\,(BW)^4 \\ & -0.540132\,exp(-09)\,(BW)^5 \\ & -0.728875\,exp(-13)\,(BW)^6 \end{aligned}$$

for BODYWEIGHTS OVER 125 Kgs.

$$\begin{aligned} 125.1 \le BW \le 135,\ SF &= 0.5208 - 0.0012\,(BW - 125) \\ 135 \le BW \le 145,\ SF &= 0.5088 - 0.0011\,(BW - 135) \\ 145 \le BW \le 155,\ SF &= 0.4978 - 0.0010\,(BW - 145) \\ 155 \le BW \le 165,\ SF &= 0.4878 - 0.0009\,(BW - 155) \end{aligned}$$

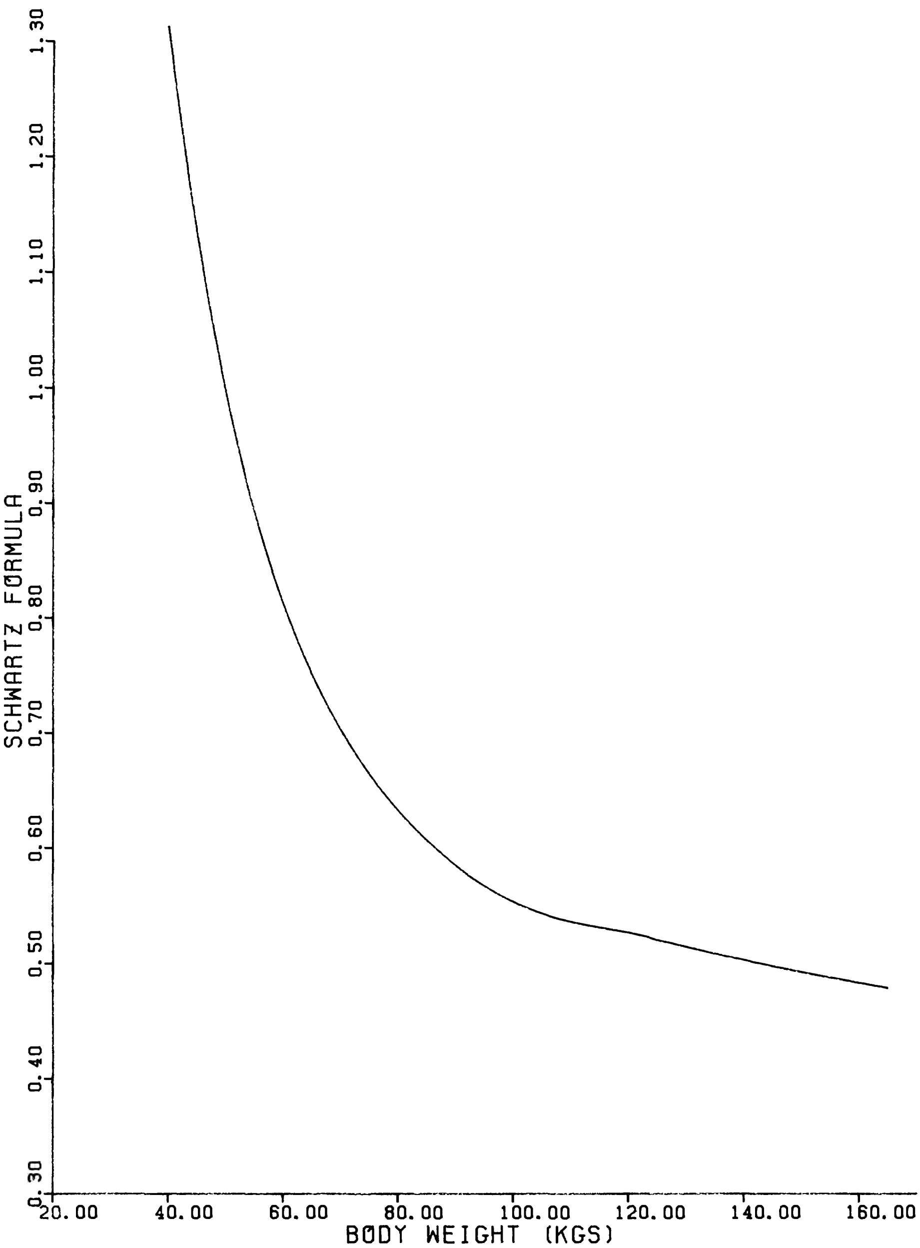
This information was made available by Dr. Lyle H. Schwartz.

A table derived from this formula is routinely used at powerlifting meets as a handicapping scheme to compensate for the widely differing body weights of the male contestants. A survey of the results of recent championship meets reveals that the formula yields sensible results.

The formula assigns each lifter a coefficient (SF) as a function of bodyweight (BW) which is then multiplied by the lifter's total [squat + bench press + deadlift (weight is lifted waist high)].

---

*The published formula incorrectly had a plus sign for the coefficient of $(BW)^3$.



Plot by Douglas Dunham

The lifter with the greatest product is awarded the best lifter trophy.

A different formula is used for women. To compensate male lifters over the age of 40 there is an additional factor which can be used in the product (e.g., Age 50 = 1.173).

Editorial comment. Isn't it curious that the formula is a 6th degree polynomial for a large portion of its domain, and piecewise linear in the rest? It is perhaps even more curious that the graph of the sextic part looks very much like an uncomplicated exponential.

Can you pass this simple true or false test on structured programming?

1 True BASIC™ is a true, structured language.

☑ True. ☐ False.

True BASIC offers a full selection of control structures like SELECT CASE, nested IF-THEN-ELSE IF, and DO-LOOP, and external procedures which can be compiled into libraries, making True BASIC faster, easier, and more flexible.

2 True BASIC has a wide range of powerful features.

☑ True. ☐ False.

True BASIC has a complete matrix algebra package and the best graphics ever in a higher level language. And there are optional libraries for things like sorting and searching and 3-D graphics.

3 Programs in True BASIC are fully portable.

☑ True. ☐ False.

Programs written in True BASIC will run on any computer which runs True BASIC, good news for anyone who uses more than one kind of personal computer.

4 *Classroom Computer Learning* named True BASIC one of the "Outstanding Products of 1985."

☑ True. ☐ False.

Not only that, but *Byte* magazine called it "Superior to Microsoft™ BASIC." *PC* magazine said it was "the easiest to learn of all the BASICS. . . ." And finally, *Electronic Learning* concluded "Good graphics have never been easier. . . ."

100%

5 Kemeny and Kurtz can also help teach students mathematics.

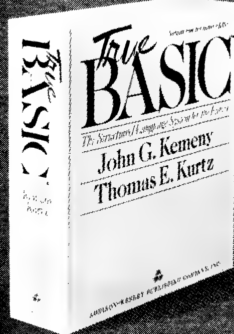
☑ True. ☐ False.

True–with the Kemeny and Kurtz Mathematics Series, including programs on algebra, trigonometry, pre-calculus, calculus, probability theory, statistics, and discrete mathematics. They're reasonably priced at only $49.95 per single copy with quantity discounts available.

O.K. You've made me a True believer.

Whether you're programming for your own applications, teaching others, or developing products to go to market, you should consider the advantages of True BASIC. After all, the true test of a good product is adoption. And True BASIC has passed that test. High schools, universities, and corporations around the world have chosen it. It's a flexible, powerful structured programming language that you can depend on, 100 percent. True BASIC is available for the IBM™ PC and compatibles, Apple Macintosh,™ and Commodore Amiga.™

To order call 1-800-TR-BASIC.
In New Hampshire, call **(603) 643-3882.**
Or send the coupon below.



True BASIC inc.
South Main Street
NH 03755

BASIC information necessary to pass this test: John Kemeny and Tom Kurtz invented the original BASIC programming language. Now they're back with an even better version: a flexible, easy-to-use structured programming language they call True BASIC.™

System requirements: IBM PC 192K, DOS 2.0+; Macintosh 128K; 512K recommended. Amiga 512K

30 Day, No-Risk Trial

Act now to take advantage of this special offer. Order today. The program costs only $149.90. Use it for 30 days, and if you're not completely satisfied, send it back for a refund.

I want to order the True BASIC Language.

My computer is: ☐ IBM PC/compatible. ☐ Apple Macintosh. ☐ Commodore Amiga. ☐ Send me True BASIC product information.

Name ______

Title ______

Company/University ______

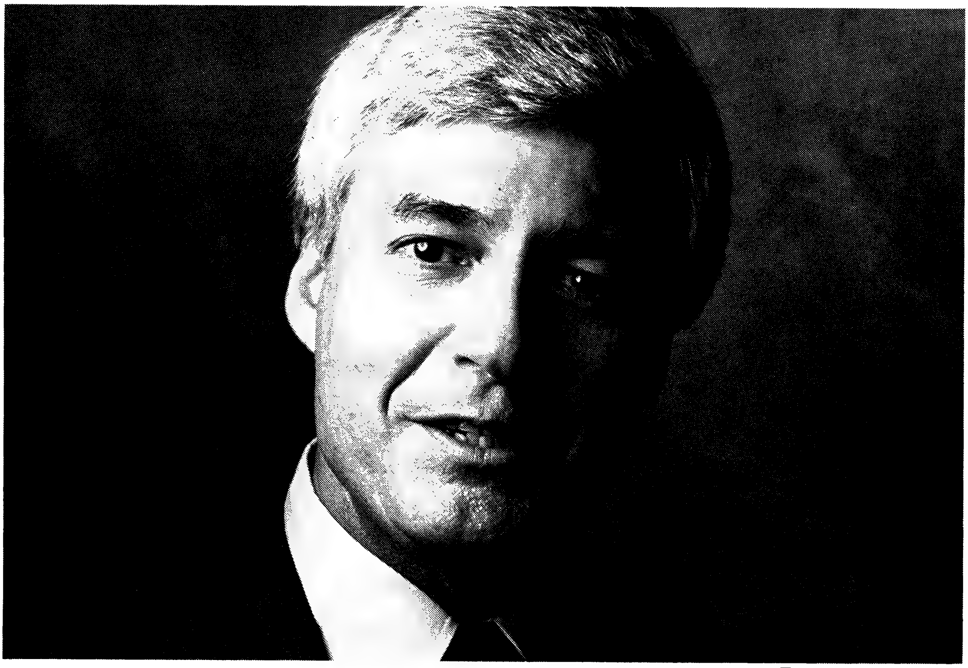
Address ______

City, State, ZIP ______

Telephone ______

Credit Card, Number, Expiration Date ______

AMM-8



JOIN MAA TODAY!

Send this application to: MAA Membership Department
1529 18th Street, N.W.
Washington, D.C. 20036

Name ______________________________

Mailing Address ______________________________

______________________________ Zip ______________

Employer/School ______________________________

Position (Rank) ______________________________

Employer's City/State ______________________________

Highest Earned Degree ______________ Year Degree Earned ______________

Institution Awarding Degree ______________________________

Month/Year of Birth ______________________________

Have you been a member of the MAA before? ☐ **Yes** ☐ **No**

Payment Enclosed $______________ *U.S. Funds Only* (see below)

Your membership may begin on July 1 or January 1. You may elect to receive any of the listed combinations of the three MAA journals:

THE AMERICAN MATHEMATICAL MONTHLY **(M)**
MATHEMATICS MAGAZINE **(G)**
THE COLLEGE MATHEMATICS JOURNAL **(J)**

Please circle the appropriate box corresponding to your initial membership period and the selection of journals you wish to receive. (Rates guaranteed for the indicated periods only.) Subscription prices are included with dues (see **).

	M	**G**	**J**	**M + G**	**M + J**	**G + J**	**M + G + J**
Student Membership							
1 year (Jan.-Dec. 1986)	$24	$18	$20	$30	$32	$26	$38
1½ year (July 1986-Dec. 1987)	$36	$27	$30	$45	$48	$39	$57
Regular Membership							
1 year (Jan.-Dec. 1986)	$40	$28	$31	$52	$55	$43	$67
1½ year (July 1986-Dec. 1987)	$60	$42	$46.50	$78	$82.50	$64.50	$100.50

*Student Membership available to high school and undergraduate students and to students regularly enrolled in graduate study at least half time. Enclose letter from school or department official confirming status. Student rates apply to unemployed persons under certain circumstances.

**Annual dues include subscription prices as follows: Reg. Memb. $24 (M), $15 (J), $12 (G), $1 (Focus). Std. Memb. $12 (M), $8 (J), $6 (G), $1 (Focus).

METHOD OF PAYMENT ☐ Check enclosed ☐ VISA ☐ MasterCard ☐ Bill me

Card Number ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐ Expiration Date ☐☐☐☐ *(month/year)*

Interbank Number ☐☐☐☐ *(MasterCard only—located above name on card)*

SIGNATURE ______________________________

MAA STUDIES IN MATHEMATICS

Studies in Computer Science

Edited by Seymour Pollack
408 pp. Hardbound List: $33.50 MAA Member:.$25.00

STUDIES IN COMPUTER SCIENCE continues the tradition of excellence of the respected series MAA Studies in Mathematics. Written by computer scientists for mathematicians, the book presents a readable and balanced discussion of the role of mathematics in computer science, and the contributions of computing to mathematics.

Opening with an historical overview of the development of computer science, the narrative continues with three articles on the nature of computer programs and the programming process, a discussion of computational complexity, and a concluding group of articles on the impact of computer science on artifical intelligence, numerical analysis, statistics, and simulation.

In the introduction the editor expresses his hope that the book will provide *"an interesting look at an explosive field in the process of becoming."* What an understatement! The process is well advanced; the impact of computer science on mathematics and its teaching is already immense. And this book is not just interesting, it is exciting and important as background reading for anyone interested in the interactions between mathematics and computer science. Teachers will find that it enriches their teaching and students will find it rewarding supplementary reading. Order your copy today!

Order From:
THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington

Carus Mathematical Monograph #21

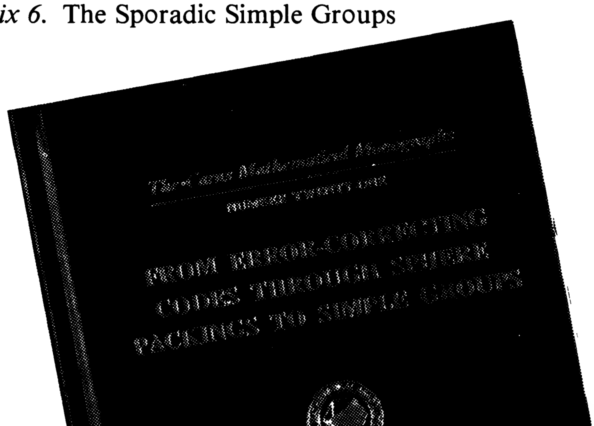
From Error Correcting Codes through Sphere Packings to Simple Groups,

by Thomas M. Thompson
224 pp. Hardbound
List: \$24.00 MAA Member: \$18.50

Two of the most fascinating problems to challenge mathematicians in recent years concern the construction of data transmission codes that can correct errors introduced by static and the search for efficient ways to pack ping-pong balls into a box. Can one design the best error-correcting codes? Can one find the most efficient sphere packing?

Therein lies a fascinating story which is told with great skill and clarity in this important addition to the Carus Mathematical Monograph series. The author has packed (sic) into 175 pages all of the basic mathematical ideas of this saga woven into a gripping historical account of the journey from error-correcting codes to sphere packings to simple groups.

Table of Contents



THE AMERICAN
MATHEMATICAL MONTHLY

Volume 93, Number 8 October 1986

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. (This is necessary for indexing purposes.) Send all main articles, letters to the editor, and editorial correspondence to the editor-elect: HERBERT S. WILF, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Contributions to the other four special departments (Notes, Teaching of Mathematics, Unsolved Problems, and Problems and Solutions) should be sent to the appropriate one of the associate editors, as follows:

Notes: CAROL G. CRAWFORD, Department of Mathematics, U.S. Naval Academy, Annapolis, MD 21402, *or* RICHARD LIBERA, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716, *or* ANITA E. SOLOW, Department of Mathematics, Grinnell College, Grinnell, IA 50112.

Teaching of Mathematics: JOAN P. HUTCHINSON *or* STAN WAGON, Department of Mathematics, Smith College, Northampton, MA 01063.

Unsolved Problems: RICHARD GUY, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.

Proposed problems and solutions, both elementary and advanced: PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801.

Reprint Permission: A. B. WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Back Issues: P. and H. BLISS CO., Middletown, CT 06457.

Microfilm Editions: University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

Editorial Assistants: Fanny Hailpern, Mary P. Jackson.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT, ten times a year: January, February, March, April, May, June-July, August-September, October, November and December.

The annual subscription price for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is $24 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $24. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $70 per year.

Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

PRINTED IN THE UNITED STATES OF AMERICA

A NEW PROOF OF ERDŐS'S THEOREM ON MONOTONE MULTIPLICATIVE FUNCTIONS

EVERETT HOWE*
Department of Mathematics, California Institute of Technology, Pasadena, CA 91125

1. Introduction. An arithmetical function f, not identically zero, is said to be *multiplicative* if

$$f(mn) = f(m)f(n) \quad \text{whenever} \quad (m, n) = 1,$$

and *completely multiplicative* if

$$f(mn) = f(m)f(n) \quad \text{for all } m \text{ and } n.$$

The following remarkable theorem concerning increasing multiplicative functions is due to Erdős [**2**].

THEOREM. *If f is increasing and multiplicative, then there is a constant α such that $f(n) = n^\alpha$ for all $n \geqslant 1$.*

Erdős's original proof is rather complicated, and simpler proofs have been given by Moser and Lambek [**3**], Besicovitch [**1**], and Schoenberg [**4**]. All of these proofs are either lengthy or not well motivated. This paper shows that the result is fairly easy to prove for completely multiplicative functions; the difficulty lies in showing that every increasing multiplicative function is completely multiplicative. Consequently, we split Erdős's theorem into two parts as follows:

THEOREM A. *If f is increasing and completely multiplicative, then there is a constant α such that $f(n) = n^\alpha$ for all $n \geqslant 1$.*

THEOREM B. *Every increasing multiplicative function is completely multiplicative.*

2. Proof of Theorem A. Let f be increasing and completely multiplicative. We prove Theorem A by contradiction. Assume there is no α such that $f(n) = n^\alpha$ for all $n \geqslant 1$. Then $\log f(n)/\log n$ is not constant, so there exist distinct integers $m > 1$, $n > 1$, such that

$$\frac{\log f(m)}{\log m} > \frac{\log f(n)}{\log n}.$$

Taking x to be the larger and y the smaller of these two ratios, we have

$$f(m) = m^x \quad \text{and} \quad f(n) = n^y,$$

with $x > y$.

Because $y/x < 1$ there exist integers $A \geqslant 1$ and $B \geqslant 1$ such that

$$\frac{y}{x} B \frac{\log n}{\log m} < A \leqslant B \frac{\log n}{\log m}.$$

In fact,

$$B = \left[\frac{x \log m}{(x - y)\log n}\right] + 1 \quad \text{and} \quad A = \left[B \frac{\log n}{\log m}\right]$$

satisfy the above inequalities. But then we have both

Everett Howe: I am currently completing my senior year at the California Institute of Technology. My mathematical interests lie chiefly in the area of number theory, particularly analytic number theory, but I have been spending a lot of time lately studying Julia sets. I am a 1985 Putnam Fellow, and I was recently awarded a National Science Foundation Graduate Fellowship.

*Supported by a Caltech Summer Undergraduate Research Fellowship.

$$A \log m \leqslant B \log n \quad \text{and} \quad Ax \log m > By \log n,$$

or

$$m^A \leqslant n^B \quad \text{and} \quad m^{Ax} > n^{By}.$$

However, since f is completely multiplicative,

$$f(m^A) = f(m)^A = m^{Ax} \quad \text{and} \quad f(n^B) = f(n)^B = n^{By},$$

so that $m^A \leqslant n^B$ while $f(m^A) > f(n^B)$, contradicting the fact that f is increasing. This proves Theorem A.

3. Proof of Theorem B. The proof of Theorem B is based on the following lemma.

LEMMA 1. *Given an increasing multiplicative function f and a prime p, let*

$$L = \inf_{x \not\equiv 0 \bmod p} \frac{f(x+p)}{f(x)}.$$

Then $L = 1$.

We use Lemma 1 to deduce Theorem B, and then we prove Lemma 1 in the next section.

Assume f is increasing and multiplicative. To show that f is completely multiplicative it suffices to show that for every prime p and every integer $n > 1$ we have

$$f(p^n) = f(p)^n.$$

Fix a prime p and let $y_n = f(p^n)$. We will show that $y_n = y_1^n$ or, equivalently, that

$$\frac{y_{n+1}}{y_n} = y_1 \tag{1}$$

for all n.

Consider any integer x not divisible by p. Then for any n we have

$$(px-1)p^n < xp^{n+1} < (px+1)p^n.$$

Now each of $px-1$, x, and $px+1$ is prime to p so

$$f(px-1)y_n \leqslant f(x)y_{n+1} \leqslant f(px+1)y_n,$$

because f is increasing and multiplicative. Therefore,

$$\frac{y_{n+1}}{y_n} \leqslant \inf \frac{f(px+1)}{f(x)} \leqslant \inf \frac{f(px+p^2)}{f(x)} = f(p) \inf \frac{f(x+p)}{f(x)} = y_1 L,$$

where the inf is taken over all $x \not\equiv 0 \bmod p$. Similarly, we find

$$\frac{y_{n+1}}{y_n} \geqslant y_1 U,$$

where

$$U = \sup_{\substack{x \not\equiv 0 \bmod p \\ x > p}} \frac{f(x-p)}{f(x)} = \sup_{x \not\equiv 0 \bmod p} \frac{f(x)}{f(x+p)} = \frac{1}{L}.$$

Thus we have

$$Uy_1 \leqslant \frac{y_{n+1}}{y_n} \leqslant Ly_1.$$

But $L = 1$ by Lemma 1, so $U = 1$ also, and this last inequality implies (1), which, in turn, proves Theorem B.

4. Proof of Lemma 1. Assume f is increasing and multiplicative. From the definition of L, if an integer x is not divisible by p we have $f(x+p) \geqslant Lf(x)$, and hence

$$f(x+kp) \geqslant L^k f(x)$$

for every integer $k \geqslant 0$.

Now, given any $k \geqslant 0$, we can find an integer $x > kp$ which is prime to both p and 2. Then $2x > x + kp$, and

$$f(2)f(x) = f(2x) \geqslant f(x+kp) \geqslant L^k f(x),$$

so that $L^k \leqslant f(2)$. Since k was arbitrary, we must have $L \leqslant 1$. But since f is increasing, we also have $L \geqslant 1$. These two inequalities show that $L = 1$ and prove Lemma 1.

5. Comments. Erdős's theorem implies a corresponding result for decreasing multiplicative functions with one restriction. It is easy to show that if f is decreasing and multiplicative, then either f is always positive, or else $f(1) = 1$ and $f(n) = 0$ when $n > 2$, with $0 \leqslant f(2) \leqslant 1$. Now, if f is positive and decreasing, then $1/f$ is increasing, so Erdős's theorem implies $1/f(n) = n^{\alpha}$ for some α, hence $f(n) = n^{-\alpha}$. Thus, if f is monotone and multiplicative and if $f(3) \neq 0$, then there exists a constant α such that $f(n) = n^{\alpha}$ for $n \geqslant 1$.

The author wishes to thank Professor Tom M. Apostol for suggesting this research and for helping in the preparation of the manuscript for publication.

References

1. A. S. Besicovitch, On Additive Functions of a Positive Integer, Studies in Mathematical Analysis and Related Topics, pp. 38–41, Stanford University Press, Stanford, CA, 1962.
2. P. Erdős, On the distribution function of additive functions, Ann. of Math., (2) 47 (1946) 1–20.
3. L. Moser and J. Lambek, On monotone multiplicative functions, Proc. Amer. Math. Soc., 4 (1953) 544–545.
4. I. J. Schoenberg, On two theorems of P. Erdős and A. Rényi, Illinois J. Math., 6 (1962) 53–58.

GRAPHICAL CONSTRUCTIONS OF MEANS

RICHARD P. SAVAGE, JR.
Department of Mathematics and Computer Science, Tennessee Technological University, Cookeville, TN 38505

1. Introduction. In chapter three of *Inequalities* by Hardy, Littlewood, and Pólya, it is shown that every continuous and strictly monotonic function defined on an interval can be used to assign a mean value to each set of n numbers in the interval. The mean value assigned is only a mean in a general sense: the mean is not less than the smallest of the numbers nor greater than the largest. Moreover, properties that the familiar means possess need not hold for the mean defined by an arbitrary function. In this paper we show that the graphical technique for determining a mean of two variables given by Moskovitz in [4] can be generalized to determining a mean of n variables (in the sense of Hardy, Littlewood, and Pólya) through a graphical technique in R^n. This provides some especially interesting geometric representations of many well-known means in R^3 and more generally in R^n.

Richard P. Savage, Jr.: I received my Ph. D. in 1981 from the University of Utah for work in differential geometry. My thesis advisor was Domingo Toledo. I was a faculty member first at Moorhead State University in Minnesota. Since 1982 I have been a faculty member at Tennessee Technological University where my father is also a member of the mathematics faculty. My nonmathematical interests include genealogical research, keeping up with baseball, and reading the classics.

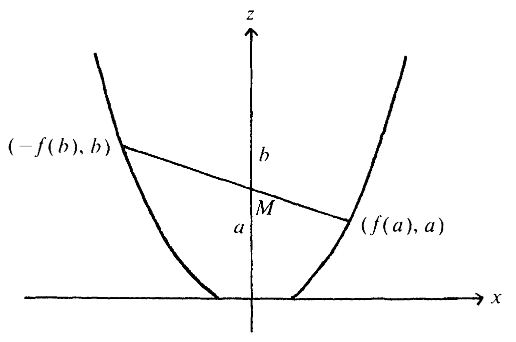


FIG. 1

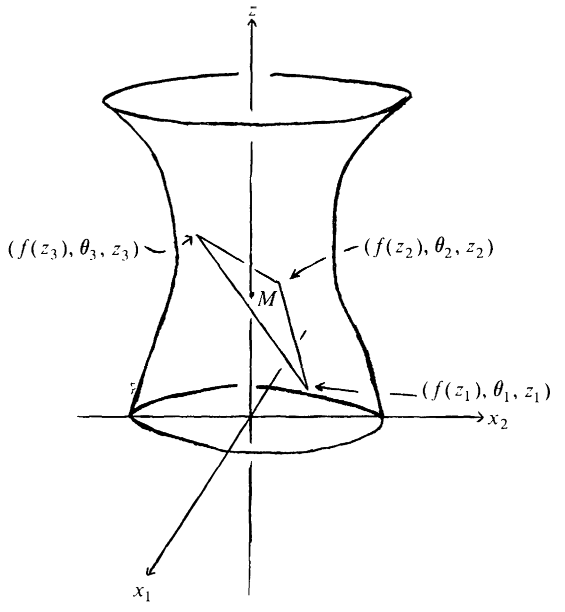


FIG. 2

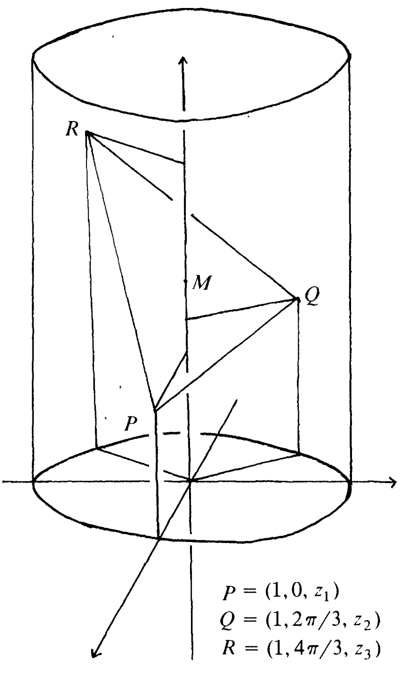


FIG. 3

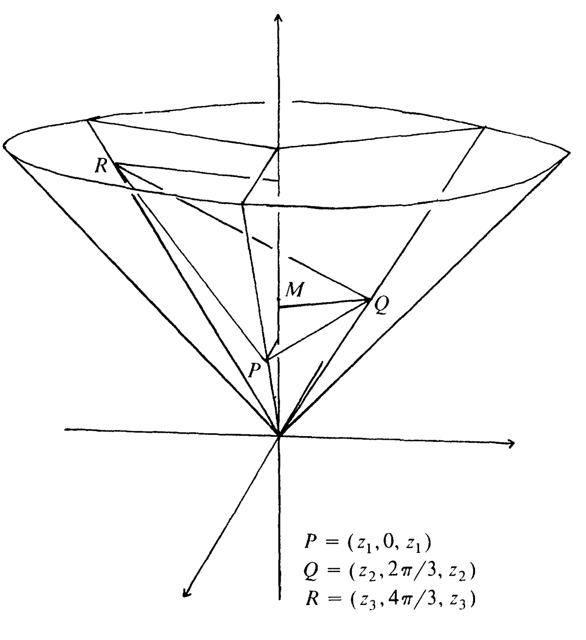


FIG. 4

If $m(z_1, z_2, \ldots, z_k)$ is a function of k variables, then we say that m is a mean if

$$\min(z_1, z_2, \ldots, z_k) \leqslant m(z_1, z_2, \ldots, z_k) \leqslant \max(z_1, z_2, \ldots, z_k) \text{ for all } z_1, z_2, \ldots, z_k.$$

A mean may satisfy other properties. A mean is symmetric if whenever σ is any permutation of $1, 2, \ldots, k$, then

$$m(z_1, z_2, \dots, z_k) = m(z_{\sigma(1)}, z_{\sigma(2)}, \dots, z_{\sigma(k)}).$$

A mean is homogeneous if

$$m(\lambda z_1, \lambda z_2, \dots, \lambda z_k) = \lambda m(z_1, z_2, \dots, z_k)$$

for all λ for which both sides are defined. There are many well-known means of positive variables. Some of these are:

Arithmetic	$A = \Sigma\lambda_i z_i$
Geometric	$G = \Pi z_i^{\lambda_i}$
Harmonic	$H = (\Sigma\lambda_i z_i^{-1})^{-1}$
Contraharmonic	$C = (\Sigma\lambda_i z_i^2)/(\Sigma\lambda_i z_i)$
Root Mean Square	$R = (\Sigma\lambda_i z_i^2)^{1/2}$
Power Mean	$M_r = (\Sigma\lambda_i z^r)^{1/r}$

where in each case $\Sigma\lambda_i = 1$ and $\lambda_i \geqslant 0$. If all λ_i are equal the means we obtain are symmetric. The power mean contains the arithmetic, harmonic, and root mean square as special cases and $\lim_{r\to 0} M_r = G$. All the means listed are homogeneous.

This paper has been motivated by the paper of Mays [3] in which is included a graphical technique due to Moskovitz [4] of associating a mean of two positive variables to a positive valued function f on the interval $(0, \infty)$. Moskovitz associates a mean to f by first constructing the graph of $x = f(z)$ in the first quadrant and then reflecting the graph across the z-axis. Then the mean m of a and b determined by f is the z-coordinate of the point where the line segment joining $(f(a), a)$ to $(-f(b), b)$ crosses the z-axis (see Fig. 1).

We can generalize this technique to means of three variables. Let f be a positive valued function defined on a subset of $(0, \infty)$ and revolve the graph of f about the z-axis to form a surface given by $[f(z)]^2 = x_1^2 + x_2^2$. Let θ_1, θ_2, and θ_3 be given and find the points on the surface whose cylindrical coordinates are $(f(z_i), \theta_i, z_i)$. Let m be the z-coordinate of the point of intersection of the plane determined by these three points with the z-axis. We show in Section 3 that with certain restrictions, m is a mean value of z_1, z_2, and z_3 (see Fig. 2).

For example, take $\theta_1 = 0$, $\theta_2 = 2\pi/3$, and $\theta_3 = 4\pi/3$. As will be shown later, letting $f(z) = 1$ we get the arithmetic mean; $f(z) = z$ yields the harmonic mean; and $f(z) = 1/z$ yields the contraharmonic mean (see Figs. 3, 4, and 5). As a final example, let $f(z) = 1$, $\theta_1 = 0$, $\theta_2 = \pi/2$, and $\theta_3 = \pi$. Then $m = (z_1 + z_2)/2$ (see Fig. 6). All coordinates in the figures are cylindrical coordinates and m denotes the mean.

There are similar geometric interpretations for means of k variables. First form the hypersurface of revolution

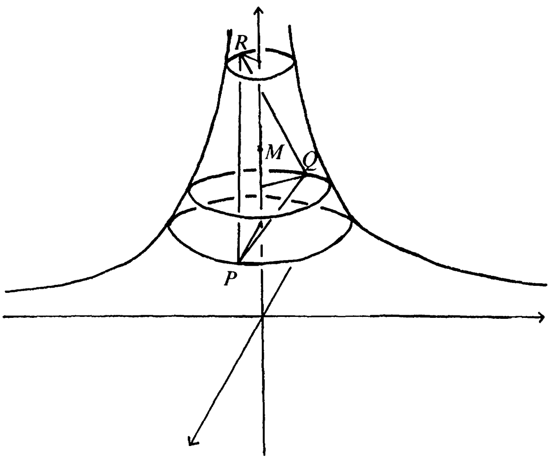
$$[f(z)]^2 = x_1^2 + \cdots + x_{k-1}^2.$$

Next locate k points on the hypersurface as described in the examples and find the intersection of the hyperplane determined by these k points with the z-axis. Note that in the definition that follows we use rectangular coordinates for the points on the hypersurface so that in the case $k = 3$ the notation $(f(z), \theta, z)$ is now replaced by $(f(z)\cos\theta, f(z)\sin\theta, z)$.

DEFINITION. Let f be a positive valued function on a subset of $(0, \infty)$ and let $v_1, v_2, \dots, v_k$ be unit vectors in R^{k-1}. Let $(x_1, x_2, \dots, x_{k-1}, z)$ be rectangular coordinates in R^k and let $v_i = (x_{i1}, x_{i2}, \dots, x_{ik-1})$. Now define $M_{f, v_1, v_2, \dots, v_k}(z_1, z_2, \dots, z_k)$, which we abbreviate as M or M_f when no confusion is possible, to be the intersection of the hyperplane determined by the points

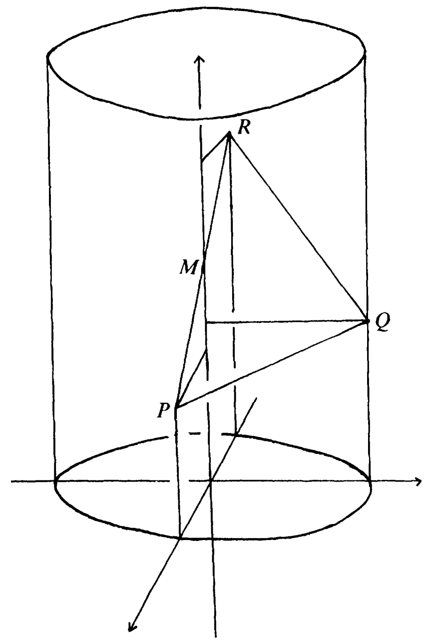
$$P_i = (f(z_i)x_{i1}, f(z_i)x_{i2}, \dots, f(z_i)x_{ik-1}, z_i), \quad i = 1, 2, \dots, k$$

with the z-axis provided that the hyperplane exists and its intersection with the z-axis is unique.



$P = (1/z_1, 0, z_1)$, $Q = (1/z_2, 2\pi/3, z_2)$, $R = (1/z_3, 4\pi/3, z_3)$

FIG. 5



$P = (1, 0, z_1)$ $Q = (1, \pi/2, z_2)$, $R = (1, \pi, z_3)$

FIG. 6

It will be seen that with certain restrictions M provides a mean of $z_1, z_2, \ldots, z_k$.

2. The means M_f. In this section we determine the conditions under which M_f determines a mean and under what further conditions M will be symmetric or homogeneous. We will also see which of the familiar means can be constructed in this way.

For M to be defined the hyperplane determined by the P_i must be unique and not contain or be parallel to the z-axis. If H is any hyperplane containing the P_i and $(x_1, x_2, \ldots, x_{k-1}, z)$ is a point of H, then

$$\begin{vmatrix} f(z_1)x_{11} - x_1 & \ldots & f(z_1)x_{1k-1} - x_{k-1} & z_1 - z \\ \vdots & & \vdots & \vdots \\ f(z_k)x_{k1} - x_1 & \ldots & f(z_k)x_{kk-1} - x_{k-1} & z_k - z \end{vmatrix} = 0, \tag{1}$$

since the row vectors of this determinant would then be linearly dependent. Now if H is not unique, every point would satisfy this equation and hence H is unique if the coefficient of some x_i or z is nonzero. The condition that H not contain or be parallel to the z-axis is equivalent to the coefficient of z being nonzero. Hence we must have

$$\begin{vmatrix} f(z_1)x_{11} & \ldots & f(z_1)x_{1k-1} & 1 \\ \vdots & & \vdots & \vdots \\ f(z_k)x_{k1} & \ldots & f(z_k)x_{kk-1} & 1 \end{vmatrix} \neq 0. \tag{2}$$

Let $A = [x_{ij}]_{k \times k-1}$, let A_i denote the submatrix of A formed by deleting the ith row, and let $D_i = (-1)^i \det A_i$. The first theorem gives conditions under which M is a mean and further conditions under which it is symmetric. Note in particular that the conditions depend only on the vectors v_i and not on f.

THEOREM 1. *M is a mean if and only if all $D_i \geqslant 0$ (or all $D_i \leqslant 0$) and some $D_i \neq 0$.*

Furthermore, the mean is symmetric if and only if all D_i are equal.

Proof. First suppose that all the D_i have the same sign and some $D_i \neq 0$. Note that by expanding the determinant in (2) by the last column we get $(-1)^k \sum_i D_i \prod_{j \neq i} f(z_j)$ and so the condition that all D_i have the same sign implies this determinant is nonzero as long as some $D_i \neq 0$. Hence H exists and is unique. Then the equation of H is given by (1) and setting $x_1 = x_2 = \cdots = x_{k-1} = 0$ and $z = M$ we can solve equation (1) for M getting

$$M = \frac{\begin{vmatrix} f(z_1)x_{11} & \dots & f(z_1)x_{1k-1} & z_1 \\ \vdots & & \vdots & \vdots \\ f(z_k)x_{k1} & \dots & f(z_k)x_{kk-1} & z_k \end{vmatrix}}{\begin{vmatrix} f(z_1)x_{11} & \dots & f(z_1)x_{1k-1} & 1 \\ \vdots & & \vdots & \vdots \\ f(z_k)x_{k1} & \dots & f(z_k)x_{kk-1} & 1 \end{vmatrix}}$$

$$= \frac{\sum\limits_i z_i D_i \prod\limits_{j \neq i} f(z_j)}{\sum\limits_i D_i \prod\limits_{j \neq i} f(z_j)},$$

or

$$(3) \qquad M = \sum_i \frac{\dfrac{D_i}{f(z_i)}}{\sum\limits_n \dfrac{D_n}{f(z_n)}} z_i .$$

Then the coefficients of the z_i are all nonnegative and sum to 1, so M clearly satisfies the definition of mean given in Section 1. Further, if all the D_i are equal we may divide them out getting

$$(4) \qquad M = \frac{\sum \dfrac{z_i}{f(z_i)}}{\sum \dfrac{1}{f(z_i)}},$$

and so M is clearly symmetric.

Conversely, if all $D_i = 0$ H is not defined. If there are D_i with different signs, assume without loss of generality that D_1 and D_2 differ in sign and that the sign of D_2 agrees with the sign of $\sum D_n / f(z_n)$. Now calculating $M(z_1, z_2, z_2, \dots, z_2)$ by formula (3) we get that the sum collapses to $\alpha z_1 + (1 - \alpha) z_2$, where $\alpha < 0$. Now this expression exceeds z_1 and z_2 when $z_2 > z_1$ and hence M is not a mean.

We may write (3) as

$$(5) \qquad M = \frac{\sum \dfrac{D_i z_i}{f(z_i)}}{\sum \dfrac{D_i}{f(z_i)}}.$$

Suppose that M is symmetric, and choose $z_1, z_2, \dots, z_k$ such that $z_1 \leqslant z_i$ for all i and $z_2 \geqslant z_i$ for all i. Then equation (5) yields

$$(6) \qquad \sum \frac{D_i(z_i - M)}{f(z_i)} = 0.$$

By the assumption of symmetry we also have

$$(7) \qquad \frac{D_1(z_2 - M)}{f(z_2)} + \frac{D_2(z_1 - M)}{f(z_1)} + \sum_3 \frac{D_i(z_i - M)}{f(z_i)} = 0.$$

Subtracting equation (7) from equation (6) yields

$$\frac{D_1(z_1 - M)}{f(z_1)} + \frac{D_2(z_2 - M)}{f(z_2)} = \frac{D_1(z_2 - M)}{f(z_2)} + \frac{D_2(z_1 - M)}{f(z_1)},$$

so that

$$D_1\left[\frac{(z_1 - M)f(z_2) - (z_2 - M)f(z_1)}{f(z_1)f(z_2)}\right] = D_2\left[\frac{(z_1 - M)f(z_2) - (z_2 - M)f(z_1)}{f(z_1)f(z_2)}\right].$$

Now $(z_1 - M)f(z_2) - (z_2 - M)f(z_1) \neq 0$ since the first term is nonpositive and the second is nonnegative, and they can't both be zero if $z_2 > z_1$. Then dividing we get $D_2 = D_1$, and repeating the argument with z_j taking the role of z_2 we get $D_1 = D_j$, so all D_i are equal as claimed and the proof is complete.

There are geometric interpretations of the conditions of Theorem 1. The condition that some D_i be nonzero is equivalent to some subset of $v_1, v_2, \ldots, v_k$ being a basis of R^{k-1}, say $v_1, v_2, \ldots, v_{k-1}$. Write v_k in terms of this basis as $v_k = \Sigma\mu_j v_j$; it is easily checked that the condition D_i all have the same sign is equivalent to all $\mu_j \leqslant 0$. Hence if $k = 2$ and $v_1 = (1)$, we must have $v_2 = (-1)$. In the case $k = 3$, the terminal point of v_3 must lie on the arc of the unit circle from the terminal point of $-v_1$ to the terminal point of $-v_2$. In the case all D_i are equal, then it is easily seen that all $\mu_j = -1$. If $k = 3$, we get $v_3 = -v_1 - v_2$. Since this must be a unit vector, calculating lengths we find that the angle between v_1 and v_2 must be $2\pi/3$. Hence in the case of three variables the mean is symmetric if and only if the three vectors are symmetric about the circle, and in this event by formula (4) the mean is independent of the choice of the three symmetric vectors. In particular v_1 may be taken to be $(1, 0)$. In fact, even in the general case M_f is invariant under rotations of R^{k-1}, so v_1 could be taken as $(1, 0, \ldots, 0)$.

Following Mays [3] we can also give conditions under which the mean is homogeneous. Note that the conditions for homogeneity depend only on f and not on the vectors v_i. Theorem 2 is similar in its general direction to Theorem 84, p. 68 of [2].

THEOREM 2. *M_f is homogeneous if and only if $f(z) = kz^p$ for some p and positive k.*

Proof. From formula (5) it is clear that if f is of this form then M_f is homogeneous. Conversely, if M_f is homogeneous then

$$\lambda M_f(z, a, \ldots, a) = M_f(\lambda z, \lambda a, \ldots, \lambda a).$$

Assume $D_1 \neq 0$. Plugging into (5) this yields

$$\lambda \frac{\dfrac{D_1 z}{f(z)} + \left(\sum_2 D_i\right)\dfrac{a}{f(a)}}{\dfrac{D_1}{f(z)} + \left(\sum_2 D_i\right)\dfrac{1}{f(a)}} = \frac{\dfrac{D_1 \lambda z}{f(\lambda z)} + \left(\sum_2 D_i\right)\dfrac{\lambda a}{f(\lambda a)}}{\dfrac{D_1}{f(\lambda z)} + \left(\sum_2 D_i\right)\dfrac{1}{f(\lambda a)}}.$$

If $\sum_2 D_i = 0$, $M_f(z_1, z_2, \ldots, z_k) = z_1$, which is homogeneous. Otherwise, cross multiplying and simplifying yields

$$\frac{f(\lambda z)}{f(z)} = \frac{f(\lambda a)}{f(a)}.$$

Substituting $g(y) = f(ay)/f(a)$ we get

$$\frac{g(\lambda z/a)}{g(z/a)} = g(\lambda),$$

so that $g(\lambda\mu) = g(\lambda)g(\mu)$. Hence g is multiplicative. It is well known that if g is multiplicative, then $g(\mu) = \mu^p$ for some p. Then

$$f(z) = f(a)g(z/a) = kz^p \quad \text{for some } k > 0.$$

It is easily seen from equation (4) that if the D_i are equal, then M_f is the arithmetic mean when $f = 1$, the harmonic mean when $f(z) = z$, and the contraharmonic mean when $f(z) = 1/z$. More generally, if $f(z) = z^{1-q}$, then $M_f = (\Sigma z_i^q)/(\Sigma z_i^{q-1})$ which is a one-parameter family of means studied by Beckenbach in [1]. In the case $k = 2$, M_f is the geometric mean when $f(z) = \sqrt{z}$. However, we now show that if $k \geqslant 3$ no f can give the geometric mean. For a symmetric homogeneous mean m, if $M_f = m$ we must have

$$\frac{\dfrac{x}{f(x)} + (k-1)\dfrac{a}{f(a)}}{\dfrac{1}{f(x)} + (k-1)\dfrac{1}{f(a)}} = m(x, a, \ldots, a)$$

for all x. Solving for $f(x)$ we get

$$f(x) = \frac{f(a)(x - m(x, a, \ldots, a))}{(k-1)(m(x, a, \ldots, a) - a)}.$$

Now letting m be the geometric mean and $a = 1$ we have

$$f(x) = \frac{x - \sqrt[k]{x}}{(k-1)\left(\sqrt[k]{x} - 1\right)} f(1),$$

which is not of the form $f(x) = cx^p$ unless $k = 2$, so that for $k \geqslant 3$ there is no f for which $M_f = G$.

3. Comparison of Means. If $M_f(z_1, z_2, \ldots, z_k) \geqslant M_g(z_1, z_2, \ldots, z_k)$ for all $z_1, z_2, \ldots, z_k$, we write $M_f \geqslant M_g$. The next theorem gives a test for when one mean is greater than another.

THEOREM 3. *$M_f \geqslant M_g$ if and only if f/g is decreasing.*

Proof. By formula (5) $M_f \geqslant M_g$ if and only if

$$\sum \frac{D_i z_i}{f(z_i)} \sum \frac{D_i}{g(z_i)} \geqslant \sum \frac{D_i z_i}{g(z_i)} \sum \frac{D_i}{f(z_i)}.$$

Equivalently

$$\sum_{i,j} \frac{D_i D_j z_i}{f(z_i) g(z_j)} \geqslant \sum_{i,j} \frac{D_i D_j z_j}{f(z_i) g(z_j)},$$

or

$$\sum_{i \leqslant j} \frac{D_i D_j (z_i - z_j)}{f(z_i) g(z_j)} \geqslant \sum_{i \leqslant j} \frac{D_i D_j (z_i - z_j)}{g(z_i) f(z_j)}. \tag{8}$$

Now if f/g is not decreasing we can find z_1, z_2 with $z_1 > z_2$ and

$$\frac{f(z_1)}{g(z_1)} > \frac{f(z_2)}{g(z_2)}. \tag{9}$$

Letting say D_1, be nonzero and plugging in z_1, z_2, and $z_3 = z_4 = \cdots = z_k = z_2$ in (8), we get

$$M_f \geqslant M_g \leftrightarrow \frac{D_1(D_2 + \cdots + D_k)(z_1 - z_2)}{f(z_1)g(z_2)} \geqslant \frac{D_1(D_2 + \cdots + D_k)(z_1 - z_2)}{g(z_1)f(z_2)}$$

and this last inequality is immediately seen not to hold in view of (9).

Conversely, if f/g is decreasing we have $f(z_i)g(z_j) \leqslant g(z_i)f(z_j)$ when $z_i \geqslant z_j$ and the reverse inequality when $z_i \leqslant z_j$. Then it is immediate that

$$\frac{D_iD_j(z_i - z_j)}{f(z_i)g(z_j)} \geqslant \frac{D_iD_j(z_i - z_j)}{g(z_i)f(z_j)}$$

in all cases. Then the inequality in (8) holds so $M_f \geqslant M_g$.

For example, using the functions previously determined to yield H, G, A, and C, we get in general $H \leqslant A \leqslant C$, and in the case $k = 2$ $H \leqslant G \leqslant A \leqslant C$.

4. Examples. It is interesting to look at the symmetric means of three variables given by familiar geometric surfaces in R^3. The following table lists some examples for which formula (4) was used to compute the means.

Function	Surface of Revolution	Formula for Mean
$\sqrt{z}$	paraboloid $z = x_1^2 + x_2^2$	$P = \dfrac{\sum\sqrt{z_i}}{\sum 1/\sqrt{z_i}}$
$\sqrt{1+z^2}$	hyperboloid of one sheet $x_1^2 + x_2^2 - z^2 = 1$	$Hy_1 = \dfrac{\sum z_i(1+z_i^2)^{-\frac{1}{2}}}{\sum(1+z_i^2)^{-\frac{1}{2}}}$
$\sqrt{z^2-1}$	hyperboloid of two sheets $z^2 - x_1^2 - x_2^2 = 1$	$Hy_2 = \dfrac{\sum z_i(z_i^2-1)^{-\frac{1}{2}}}{\sum(z_i^2-1)^{-\frac{1}{2}}}$
$\sqrt{1-z^2}$	sphere $x_1^2 + x_2^2 + z^2 = 1$	$S = \dfrac{\sum z_i(1-z_i^2)^{-\frac{1}{2}}}{\sum(1-z_i^2)^{-\frac{1}{2}}}$
z	cone $x_1^2 + x_2^2 = z^2$	$H = \dfrac{3}{\sum 1/z_i}$
1	cylinder $x_1^2 + x_2^2 = 1$	$A = \dfrac{\sum z_i}{3}$

By Theorem 2 the second, third, and fourth examples are nonhomogeneous. Applying Theorem 3 we see that if all z_i lie in the interval $(0,1)$,

$$S \geqslant A \geqslant Hy_1 \geqslant P \geqslant H,$$

while if all z_i lie in $(1,\infty)$,

$$A \leqslant P \leqslant Hy_1 \leqslant H \leqslant Hy_2.$$

The result holds for the corresponding means of k variables.

Another interesting example is furnished by the one-parameter family of means determined by

$f_b(z) = (z+b)/(1+b)$. In the case $k = 3$, our corresponding surfaces are circular cones with vertex $(0,0,-b)$ and containing the point $(1,0,1)$. This determines the mean

$$M_{f_b} = \frac{\sum \frac{z_i}{z_i + b}}{\sum \frac{1}{z_i + b}}.$$

For $b = 0$ this is the harmonic mean and $\lim_{b\to\infty} M_{f_b}$ is the arithmetic mean. A similar example using the function $f(z) = (1+b)/(z+b)$ yields for $b = 0$ the contraharmonic mean and $\lim_{b\to\infty} M_{f_b}$ is the arithmetic mean.

References

1. E. F. Beckenbach, A class of mean value functions, this MONTHLY, 57 (1950) 1–6.
2. G. H. Hardy, J. E. Littlewood, and G. Pólya, Inequalities, Cambridge University Press, London, 1967.
3. M. E. Mays, Functions which parametrize means, this MONTHLY, 90 (1983) 677–683.
4. D. Moskovitz, An alignment chart for various means, this MONTHLY, 40 (1933) 592–596.

GINI MEANS

DAVID FARNSWORTH AND RICHARD ORR
Department of Mathematics, Rochester Institute of Technology, Rochester, NY 14623

Corrado Gini (1883–1965), a prolific author who founded the journal of statistics, *Metron*, defined an interesting class of means. For data $X = \langle x_i | i = 1,2,\dots,n\rangle$, the Gini family of means is defined, for parameters r and s, as

$$G(r,s;X) = \left(\Sigma x_i^{r+s}/\Sigma x_i^s\right)^{1/r} \text{ [6]}.$$

The family includes the so-called power means $P(r;X) = (\Sigma x_i^r/n)^{1/r}$ (for $s = 0$), the contraharmonic mean (for $r = s = 1$), and the self-weighting means $\Sigma x_i^{1+s}/\Sigma x_i^s$ (for $r = 1$). We call the last self-weighting because they are of the form of a weighted average $\Sigma x_i^s x_i/\Sigma x_i^s$ where the weights x_i^s are generated by the data. In Sections 1 and 2, we discuss some fundamental properties of Gini means and their relation to power means. In Sections 3 and 4, we provide a simple geometric construction for some of the Gini means including the self-weighting means for integer s. We conclude, in Section 5, with a simple construction for five power means.

1. Some properties of the Gini means. The Gini means $G(r,s;X)$ for $x_i > 0$ share some of the standard properties which are associated with averages. For example, $\min\{x_i\} \leqslant G(r,s;X) \leqslant \max\{x_i\}$, that is, they satisfy the betweenness property. Also, if all $x_i = a$, then $G(r,s;X) = a$, the identity property. $G(r,s;tX) = tG(r,s;X)$; so, they are homogeneous of degree one. The Gini means are symmetric in that, if Y is a permutation of X, then $G(r,s;X) = G(r,s;Y)$.

The Gini means may differ from ordinary averages in several respects. Most of the ordinary

---

David Farnsworth received his Ph.D. in mathematics from the University of Texas at Austin in 1967. His early work was in the mathematical side of general relativity. Beginning in the mid-1970s he became active in statistics. His special interest is the exploratory data analytic technique median polish. Most of his research has been in applied areas.

Mr. Richard Orr has attended John Carroll University, Case Institute of Technology, and State University of New York at Buffalo.

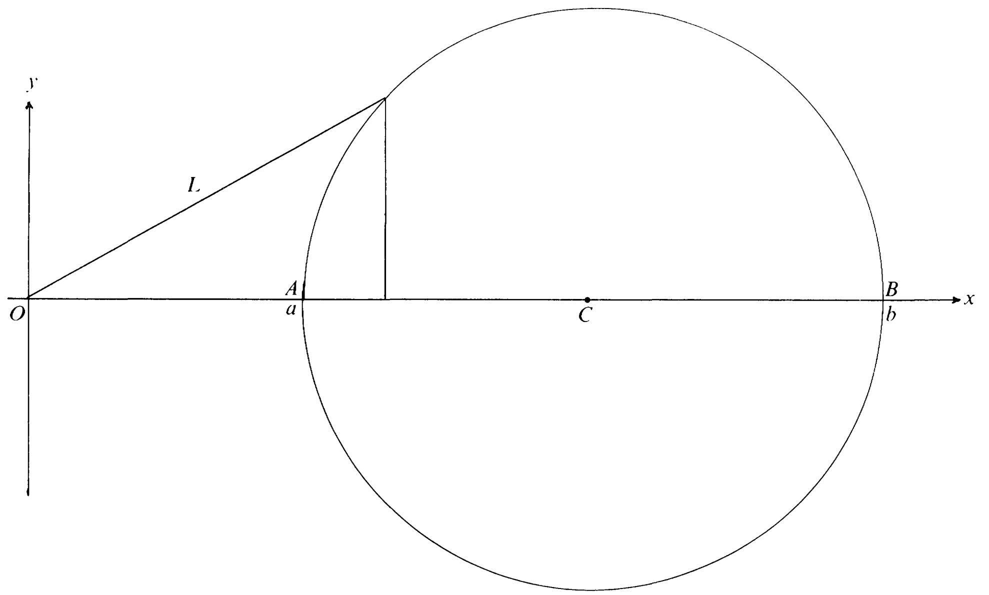


FIG. 1. $x^2 + y^2 - (a + b)x + ab = 0$.

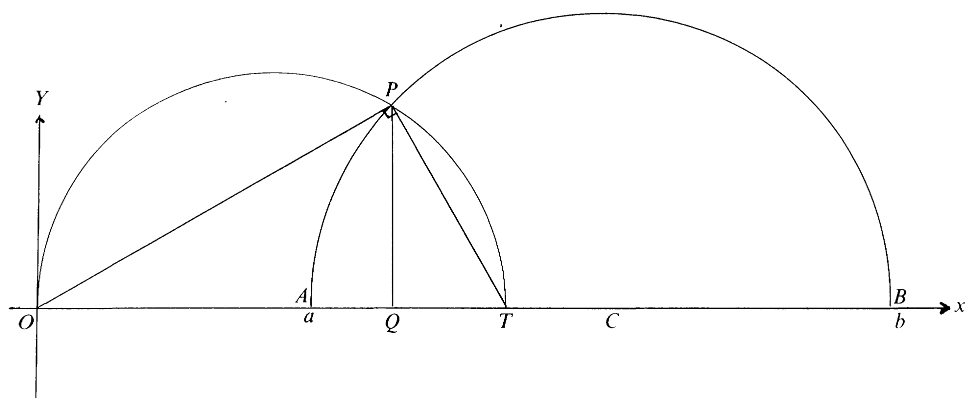


FIG. 2. Construction of T from Q and Q from T.

averages satisfy the substitution rule: if several data values are replaced by their average, the overall average is unchanged. In general, the Gini means do not satisfy this rule. For example, for the contraharmonic mean of $1, 2, 3$, we have $G(1,1;1,2,3) = 7/3$; the average for the data $1, 2$ is $G(1,1;1,2) = 5/3$, and $G(1,1;5/3,5/3,3) = 131/57 \neq 7/3$.

Commonly used averages are monotonic functions of the data values: the power means increase with each x_i, and the median is not decreasing. It is known that for any number of positive data, the Gini means with $r = 1$ are increasing for $-1 \leqslant s \leqslant 0$ and for certain parts of their domain for other s [1].

For the contraharmonic mean $(a^2 + b^2)/(a + b)$, let b be fixed and let a vary. For $0 < a \leqslant b$, one has a convex function of a. As a increases from 0 to $(\sqrt{2} - 1)b$, the mean decreases from b to $2(\sqrt{2} - 1)b \approx 0.828\, b$. Then, as a increases from $(\sqrt{2} - 1)b$ to b, the mean increases to b. Thus, not only is the mean not monotonic, but also the Gini means can be very insensitive to certain data.

2. Relationship with the power means. The classical power means are $P(r;\ X) = (\Sigma x_i^r/n)^{1/r}$ [**8**]. The power means are monotonic in each variable. As mentioned, the Gini means include the power means $G(r,0;\ X) = P(r;\ X)$. The lack of monotonicity exhibited in Section 1 shows that not all Gini means are power means.

It is well known that the power means with $r = 0$ are the unique numbers c which minimize $\Sigma(x_i^r - c^r)^2$ [**1**]. That the Gini means $G(r, s;\ X)$ are the unique numbers c which minimize $\Sigma x_i^s(x_i^r - c^r)^2$ can be shown by differentiating with respect to c.

The subset of Gini means with $r = 1$, that is, our $G(1, s;\ X)$, includes for $s = 0$ the arithmetic mean $P(1;\ X)$ and for $s = -1$ the harmonic mean $P(-1;\ X)$. These two are the only means common to the two classes $G(1, s;\ X)$ and $P(r;\ X)$ [7], [**11**]. For data sets consisting of just two values, $G(1, -\frac{1}{2};\ a, b) = P(0;\ a, b)$, the geometric mean which is defined as $\lim_{r\to 0} P(r;\ X)$.

In the next section, we provide a geometric construction for $G(1, s;\ X)$ with s any integer.

3. A geometric construction for some Gini means. We provide a geometric construction for the self-weighting Gini means ($r = 1$)

$$G(1, s;\ a, b) = (a^{s+1} + b^{s+1})/(a^s + b^s),$$

where s is any integer, by extending a construction of Schoenberg [**14**]. For fixed a, b, let us simply denote this by $G(s)$. Accordingly, let a, b be unequal positive real numbers with $b > a$ ($a = b$ is a trivial case) and consider the circle on AB as diameter where A and B are the points $(a, 0)$ and $(b, 0)$ in a cartesian plane (Fig. 1). The circle has center $C(\frac{1}{2}(a + b), 0)$, radius $\frac{1}{2}(b - a)$, and its equation is

$$\left(x - \frac{1}{2}(a + b)\right)^2 + y^2 = \left(\frac{1}{2}(b - a)\right)^2$$

or

$$x^2 + y^2 - (a + b)x + ab = 0.$$

For the point (x, y) on the circle, the length L of the hypotenuse of the triangle with vertices $(0, 0)$, (x, y), and $(x, 0)$ is given by

$$L^2 = x^2 + y^2 = (a + b)x - ab.$$

Referring to Fig. 2, suppose, at any point Q on the x-axis between A and B, a line perpendicular to the x-axis is constructed to intersect the circle at P, and the line through P, perpendicular to OP, is constructed to intersect the x-axis at a new point T. In semicircle APB, PB is perpendicular to PA. Since the inclination of OP is not as great as that of AP, then PT, which is perpendicular to OP, must place T on the x-axis between A and B. Conversely, for any point T on the x-axis between A and B, the circle with diameter OT can be constructed to intersect the original circle at P, from which a perpendicular to the x-axis gives Q. Hence, from the position of either Q or T, the other can be constructed.

From the proportional sides of similar right triangles OPQ and OPT, we have

$$OQ \cdot OT = OP^2. \tag{1}$$

For a point $P(x, y)$ on the circle with diameter AB we have

$$OP^2 = x^2 + y^2 = (a + b)x - ab = (a + b)OQ - ab,$$

and we obtain

$$OQ \cdot OT = (a + b)OQ - ab.$$

Now, if OQ happened to have the value $G(s)$, it would follow that

$$OT = (a + b) - \frac{ab}{OQ} = a + b - ab\left(\frac{a^s + b^s}{a^{s+1} + b^{s+1}}\right)$$

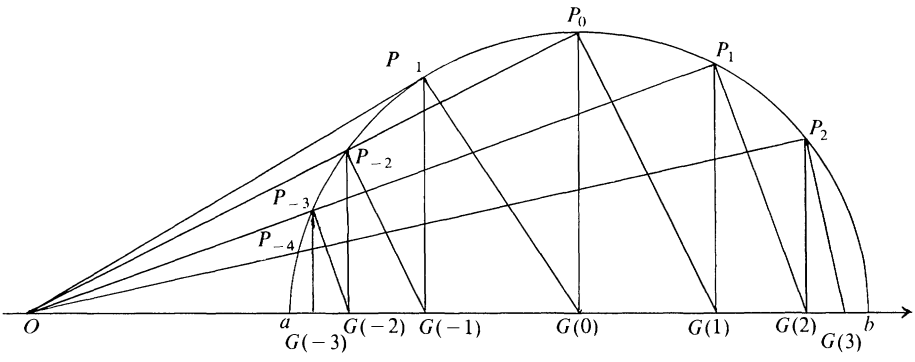


FIG. 3. Iterating to obtain the Gini means $G(s)$ for s between -3 and 3.

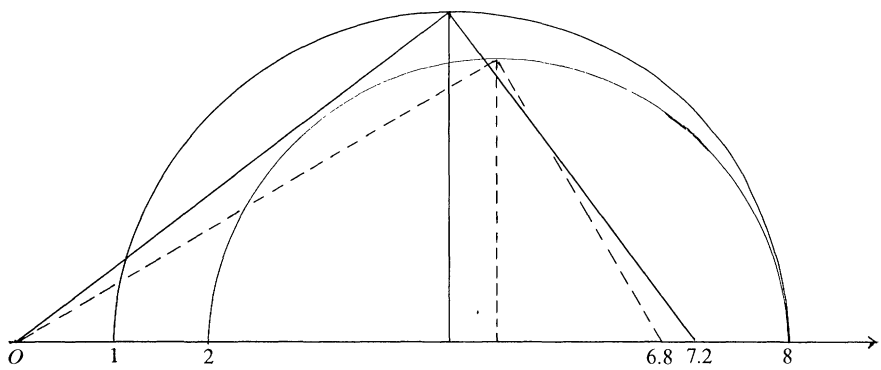


FIG. 4. Increasing a data value may decrease a Gini mean.

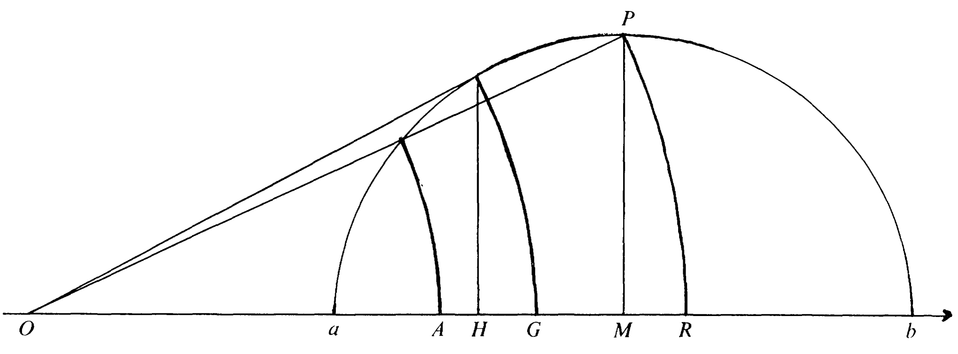


FIG. 5. The power means $((a^r + b^r)/2)^{1/r}$ along the horizontal axis for $r = -2, -1, 0, 1, 2$.

$$= \frac{a^{s+2} + b^{s+2}}{a^{s+1} + b^{s+1}} = G(s+1).$$

Conversely, for $OT = G(s)$, we obtain $OQ = G(s-1)$.

Observe that for Q at the center C, we have $OQ = \frac{1}{2}(a+b) = G(0)$. Therefore, starting with Q at C and iterating our construction for T, we obtain points on the x-axis whose abscissae are $G(1), G(2), G(3), \dots$; similarly, beginning with T at C and iterating the reverse construction for the point Q, we obtain points having abscissae $G(-1), G(-2), G(-3), \dots$. See Fig. 3.

For an integer s, then, we have that $G(s) < G(s+1)$ and that $G(s)$ is a value betwen a and

b, that is, $G(s)$ has the betweenness property. More generally for data set X, the class of Gini means $G(1, s; X)$ is an increasing function of s. This can also be shown using calculus methods **[1]**. Note, in addition, that the class of power means $P(r; X) = G(r, 0; X)$ is an increasing function of r **[8]**.

The symmetry of Fig. 3 leads to an unexpected relationship; it suggests $G(s) + G(-s) = 2G(0)$ which quickly can be confirmed algebraically.

The construction above also yields interesting results when applied to other choices for OQ. For example, if $OQ = ta + (1 - t)b$, then

$$OT = \frac{ta^2 + (1-t)b^2}{ta + (1-t)b},$$

a weighted Gini mean.

4. Further constructions. In Fig. 3 the distances OP_s, which we denote by $L(s)$, are Gini means, too. Equation (1) reveals that $L(s)$ is the geometric mean of $G(s)$ and $G(s+1)$, which gives

$$L(s) = G(2, s; a, b) = \left(\frac{a^{s+2} + b^{s+2}}{a^s + b^s}\right)^{1/2}.$$

Because the geometric mean satisfies the betweenness property, $G(s)$ and $L(s)$ interlace, that is,

$$(2) \qquad \cdots < L(s-1) < G(s) < L(s) < G(s+1) < L(s+1) < \cdots$$

and $a < L(s) < b$.

The lack of monotonicity can be displayed by the construction. In Fig. 4 it is shown that $G(1,1;2,8) = 6.8 < 7.2 \approx G(1,1;1,8)$. Increasing the first data value from 1 to 2 decreases the mean in this example.

5. An interesting special case. Five power means are produced by the Gini-mean construction. As in Section 3, $G(0) = M$, the arithmetic mean, and $G(-1) = H$, the harmonic mean. As in Section 4, $L(0) = R$, the root-mean-square, $L(-1) = G$, the geometric mean, and $L(-2) = A$, the harmonic-root-mean, $(\Sigma x_i^{-2}/n)^{-1/2}$. The interlacing property (set $s = -1$ in (2)) gives $A < H < G < M < R$. Fig. 5 displays all five means. For constructions with a similar goal, see **[2]**, **[3]**, **[4]**, **[5]**, **[9]**, **[10]**, **[12]**, **[13]**, **[14]**, and **[15]**.

References

1. E. F. Beckenbach, A class of mean value functions, this MONTHLY, 57 (1950) 1–6.

2. E. F. Beckenbach and R. Bellman, An Introduction to Inequalities, New Mathematical Library No. 3, Mathematical Association of America, Washington, DC, 1961.

3. J. L. Ercolano, Geometric interpretations of some classical inequalities, Math. Mag., 45 (1972) 226.

4. ______, Remarks on the neglected mean, Math. Teacher, 66 (1973) 253–255.

5. J. Garfunkel and B. Plotkin, Using geometry to prove algebraic inequalities, Math. Teacher, 59 (1966) 730–734.

6. C. Gini, Di una formula comprensiva delle medie, Metron, 13 (1938) 3–22.

7. H. W. Gould and M. E. Mays, Series expansions of means, J. Math. Anal. Appl., 101 (1984) 611–621.

8. G. H. Hardy, J. E. Littlewood, and G. Pólya, Inequalities, Cambridge University Press, London, 1967, Chapter 3.

9. L. Hoehn, A geometric interpretation of the weighted mean, College Math. J., 15 (1984) 135–139.

10. K. Iles and L. J. Wilson, A further neglected mean, Math. Teacher, 70 (1977) 27–28.

11. D. H. Lehmer, On the compounding of certain means, J. Math. Anal. Appl., 36 (1971) 183–200.

12. E. Maor, A mathematician's repertoire of means, Math. Teacher, 70 (1977) 20–25.

13. A. Schild, Geometry of the means, Math. Teacher, 67 (1974) 262–263.

14. I. J. Schoenberg, Mathematical Time Exposures, Mathematical Association of America, Washington, DC, 1982, Chapter 12.

15. J. C. Slay and J. L. Solomon, A mean generating function, Two-Year College Math. J., 12 (1981) 27–29.

A great man with a small name. (See p. 630.)

so that $\|x - Tx\| = \lim_{p\to\infty}\|x_{N_p} - Tx_{N_p}\| = 0$.

We turn now to the proof of Markov's theorem. Define a convex compact subset K of $\mathbb{R}^n$ as follows:

$$K = \left\{ x = (x_1, \dots, x_n) \in \mathbb{R}^n \colon x_i \geqslant 0 \quad \text{for all } i,\ \sum_{i=1}^{n} x_i = 1 \right\},$$

and define a linear mapping $T\colon \mathbb{R}^n \to \mathbb{R}^n$ as follows:

$$Tx = \left(\sum_{j=1}^{n} p_{1j}x_j, \sum_{j=1}^{n} p_{2j}x_j, \dots, \sum_{j=1}^{n} p_{nj}x_j \right).$$

The mapping T is continuous because

$$\|Tx\| = \max_{1\leqslant i\leqslant n}\left| \sum_{j=1}^{n} p_{ij}x_j \right| \leqslant \|x\| \max_{1\leqslant i\leqslant n} \sum_{j=1}^{n} p_{ij},$$

so that letting $c = \max\limits_{1\leqslant i\leqslant n} \sum\limits_{j=1}^{n} p_{ij}$ we have

$$\|Tx\| \leqslant c\|x\|.$$

CLAIM: *The restriction of T to K is an affine continuous mapping from K to K.*

Since T is linear continuous on $\mathbb{R}^n$, its restriction to K is affine continuous. We need only prove that $x \in K$ implies that $Tx \in K$. Let $x \in K$, $x = (x_1, \dots, x_n)$, then by definition:

$$Tx = \left(\sum_{j=1}^{n} p_{1j}x_j, \dots, \sum_{j=1}^{n} p_{nj}x_j \right).$$

For each i we have $\sum\limits_{j=1}^{n} p_{ij}x_j \geqslant 0$. Furthermore

$$\sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}x_j = \sum_{j=1}^{n} x_j \sum_{i=1}^{n} p_{ij} = \sum_{j=1}^{n} x_j = 1$$

and so $Tx \in K$, proving our claim.

Applying the Markov-Kakutani Theorem we obtain the existence of some $\tilde{x} = (\tilde{x}_1, \dots, \tilde{x}_n) \in K$ with $T\tilde{x} = \tilde{x}$. This means

$$\sum_{j=1}^{n} \tilde{x}_j = 1 \quad \text{and} \quad \sum_{j=1}^{n} p_{ij}\tilde{x}_j = \tilde{x}_i \quad \text{for } i = 1, 2, \dots, n,$$

which proves Markov's Theorem.

References

1. N. Dunford and J. Schwartz, Linear Operators, Part I, Interscience Publishers, New York, 1958.
2. W. Feller, An Introduction to Probability Theory and Its Applications, Vol. 1, 3rd ed., Wiley, New York, 1968.
3. J. Franklin, Mathematical methods of economics, this MONTHLY, 90 (1983) 229–244.

ANSWER TO PHOTO ON PAGE 608

Felix Klein. The picture was taken when he was 29 years old.

REMARKS ON THE HISTORY AND PHILOSOPHY OF MATHEMATICS

ROBERT L. LONG
Department of Mathematics, University of Florida, Gainesville, FL 32611

The history and philosophy of mathematics bear only on the periphery of the work of most mathematicians whether they are teachers or researchers. As a consequence there builds up a 'received view' of these subjects, a body of dogma which remains for the most part unarticulated. This dogma influences our work even if the influence seems small and comes from afar. The primary purpose of this essay is to bring the received dogma into view and examine it. In the process we will see that it is more influential than we supposed. Implicit in the entire undertaking is my conviction that an understanding of the history and philosophy of mathematics is interesting in itself and can contribute to our effectiveness as teachers.

Relevance. Before beginning the remarks that constitute the main body of this essay I should sketch an argument to make plausible my assertion that they can be relevant to us as teachers. My aim here is not so much to build a tight case as to suggest a possibility. I believe that the remarks themselves will strengthen the case.

George Pólya argues persuasively (in *Mathematical Discovery*, [**31**], see especially section 5 of the preface, section 14.2, and comment 14.5) that the primary aim of mathematical instruction is to teach problem solving. Obviously there is room for discussion about what constitutes a problem and the word 'problem' can take on a meaning so wide as to render it useless for thinking. But let us accept Pólya's assertion as a provisional guideline.

A problem is a question posed for consideration or solution. I usually approach a problem guided by the analogy with a journey. One must know where one is, where one wants to go, and then try to find a route. In some situations the first two elements are clear and we may even develop an algorithmic procedure for solving problems of a particular type without having to give them further thought. Solving linear equations in one unknown is an example of this sort. In other cases it requires a lot of thought about the destination before there is any hope of trying to find a route. Many proofs are of this sort; it is difficult to get a clear conception of what is to be proved. The problem of disposing of nuclear wastes is also of this sort. It appears that elementary instruction (including the college level 'service course') often assumes an overly narrow conception of a problem (first type only) and then makes things worse by going directly to the algorithm instead of allowing it to reveal itself as the labor saving device it is.

To return to the instructional problem, we believe that we understand the destination. The point of departure is not as clear. In each class we must make an effort to find out *where the students are* and then to chart a course. For the most part this means: choose what we will use to try to motivate them. A great variety of analogies are available for motivation. Their usefulness is related to some philosophical issues which I will discuss later. For now it will suffice to glance at what all too often passes for motivation.

Of course the purpose of motivation is to help propel the student along the path from where he is now to his destination. But what is his destination? Are we to ask him? We can hardly hope to tell him; after all, how does one motivate a student to achieve an end he hasn't accepted as his own? A goal and the process of working toward it are a dialectical pair. The vaguely conceived goal serves as an initial impetus to undertake a path of study. As he moves along that path the student's conception of the goal may change in accordance with what he has newly learned. The

---

Robert Long was educated at Haverford College and the University of Illinois were he earned a Ph.D. under Leon McCulloh in algebraic number theory. He enjoys teaching and studying mathematics, thinking, and playing the guitar in addition to generally being alive.

altered goal now reflects back and influences the further course of study. In most cases this is a subtle process of which the student may not be explicitly aware. We should be aware of it just as we should be aware of some reasons for learning that may not have occurred to a particular student.

But wait! Isn't this concern with motivation a tempest in a teapot? After all, for some time now everyone has known that mathematics is the language in which science describes and teaches us to manipulate the world. Computers are taking over more and more of the day to day tasks of the commercial world in which our students will earn a living. Most of these students are already convinced of the need to understand at least some mathematics.

Whenever we use one thing to motivate the study of a second we are also tacitly asserting that the first is more important than the second. This is implicit in the 'for-the-sake-of'. It would be foolish to act as though future prosperity is not or should not be important to students, but it is just as foolish to permit it to stand as the only or even the most important source of motivation. In the final analysis the conception of mathematics as a 'mere tool' is not only philosophically naive, it is debilitating to mathematics and alienating to students.

The alienation is rooted in a belief in a clearcut separation between the real world and the mathematical tools with which science describes and manipulates it. Students see that in order to function successfully in the world they must master those tools but they don't see an *essential* connection between the two. (This is very clear in the belief that reality would 'be the same' without mathematics.) Consequently they see the mathematics as a mere obstacle.

Some teachers accept this utilitarian view of mathematics and rely on it exclusively for motivation. (Sadly, this is often the case in the lower grades.) Others try to cultivate an appreciation for the beauty of mathematical results. These two views of mathematics (the utilitarian and the aesthetic) both tend to abet the belief in its isolation from the mainstream of day to day experience. I think this belief is false and unnecessary.

There is a large group of students to whom the utilitarian appeal seems crass and uninteresting but who are not so captivated by the beauty of mathematics as to want to learn it. These students can best be motivated by the teacher who has a rich understanding of the role mathematics plays in experience and is able to see its relationships to the most disparate human undertakings. It is part of the task of a philosophy of mathematics to generate such an understanding.

History. I am not concerned here with what we teach in courses on the history of mathematics nor with how or to what extent we incorporate historical remarks into our regular mathematics courses. These are issues which would require separate treatment even if, as I hope, some of the remarks which follow are relevant to them. My concern is to explicate the received view of the history of mathematics and to show its relevance.

Everyone knows that there is a regrettable ignorance of the history of mathematics. Thus it may seem misleading to speak of a 'received view'. But it is precisely the nature of the received view not to be explicit. Indeed, there has been much serious work done in the history of mathematics but it has barely influenced the mathematical community at large.

My synopsis of the 'received view': the history of mathematics is the story of what ideas emerged and what theorems were proved when and in what order. It may glance at relations to other subjects but its primary task is to show how mathematics has grown, how ideas have been synthesized and new ideas generated. On this view the reasons to study the history of mathematics are (1) to have a deeper understanding and appreciation of the results one has already learned, and (2) to realize that however beautiful and general present day mathematics is, it may develop further. A time will surely come when the theory of linear operators on Hilbert space will be routine, but that theory will retain its beauty. (At this point there should pass through the reader's mind the image of a faded photograph of a once beautiful woman; the image grows misty and evaporates.)

This history of mathematics is very different from most other histories. The differences are explicable in terms of the implicit conception of mathematics: a Euclidean system of eternal truths. The history of mathematics is the record of the successive unveilings of these truths to human intelligence. As such it is an unambiguous instance of *progress*. The most important difference between the history of mathematics and other histories is that it does not involve interpretation. The meaning of a mathematical result lies solely in its relations to the rest of mathematics. Thus, anachronism, a danger in most historical studies, is the procedure of choice in regard to mathematics. (Unless the timelessness of mathematical truths makes the use of "anachronism" improper.) The whole of mathematics, as revealed up to the present, is the exclusive standard by which the significance of a result can be legitimately assessed.

The severity of this conception no doubt accounts for the lack of widespread interest in or knowledge of the history of mathematics. However, it has been noticed in recent years that historical vignettes can add "human interest" to the study of mathematics. For example in [**33**], p. 336, we are informed that Newton was once sent to cut a hole in a barn door for cats to go in and out. Unfortunately such irrelevancies are probably counterproductive; they signal the student that the mathematics they accompany is indeed a bitter pill.

I hope that the reader objects to such an emasculated view of mathematics and its history. After all, we have provisionally agreed that an important function of mathematics is to generate methods for solving problems. This is a peculiarly human endeavor. While it is accurate to describe some animal behavior as problem solving, it appears that only man holds the problem before himself and reflects on how to solve it. Thus, to the extent that mathematics is the study or science of solving problems, it is as distinctively 'human' as literature or religion. (What this implies for the Euclidean conception remains to be unfolded.)

"But isn't it appropriate to separate mathematics proper from the essentially accidental human conditions of its discovery and use?" There is probably no point in trying to specify what this objection means by "mathematics proper". When it is pursued rigorously, the distinction which seems so clear on first glance melts into obscurity. What is more important is that if "mathematics proper" can carry a meaning in the objection then historical considerations are utterly irrelevant to understanding it.

In my view mathematics is like music in some respects; it is a creation of the human mind which generates its own standards of excellence or beauty rather than accepting them directly from other aspects of human experience. The quality of a particular solution, proof, or composition is primarily an internal affair. But a full understanding of mathematics (or music) can only come by seeing it in relation to the panorama of human possibilities. Thus the history of mathematics should not only record the development of mathematical ideas but it should also put in evidence the conceptions of mathematics which have prevailed at various times and exhibit how mathematics has participated in man's intellectual and practical life.

Now I can be more specific about the deficiencies of the received view of history.

1. It is blind to essential changes that have occurred in the role and conception of mathematics. The treatment of history in terms of evolution of important mathematical ideas (as for example in Eves [**6**] or Kramer [**23**]) needs to be supplemented by one based on a much broader view. For example, the conception of mathematics (and consequently its essential function) has changed very much from the classical Greek to the Christian and again from the Christian to the modern scientific period (since Galileo). An appreciation of these changes leads to a better understanding of the foundational crises which have dominated the landscape in philosophy of mathematics in this century. Notwithstanding certain superficial claims regarding the discovery of the incommensurability of the side and diagonal of a square, a 'crisis of the foundations' is not conceivable within the context of Greek mathematics.

When essential changes in the role of mathematics remain unnoticed it (mathematics) comes to

be thought of as highly isolated from most human activity. This belief in its isolation is operative in the public at large and in our students. It is a major obstacle to understanding mathematics.

2. The received history of mathematics hardly recognizes the numerous false starts and incorrect proofs that have always been a part of mathematics. While there is good reason not to dwell on them, the cumulative effect of ignoring them is a picture of a science marching directly and inexorably to its goal. This helps perpetuate the mathematical mystique: that good people don't make mistakes. We know it is not so but the 'knowledge' is too abstract. We 'tell' our students that everyone makes mistakes; if we had a more accurate picture of the past our telling would carry more conviction.

3. Too seldom is explicit notice taken of the fact that what we 'cover' in one or two semesters of calculus is a distillation of over 150 years work of some very brilliant people. Of course, this reflects a pedagogical decision to provide the student with needed mathematical tools as quickly as possible. (A 'decision' which is often made on the basis of habit alone, without reflection, and is not always recognized as a decision.) But we owe it to our students to remind them that mathematics was not discovered inscribed in stone but is the creation of human beings like themselves struggling with problems.

4. The role of fashion and the influence of individuals is only rarely acknowledged. There are leaders, followers, and fads in mathematics just as there are in Shakespearian criticism, jazz, and clothing. This is a human phenomenon (or perhaps an all-too-human one). To recognize it not only helps to overcome the isolation of mathematics, it also makes it possible to look into the question: What qualities distinguish the mathematical leaders?

5. The history of mathematics is conceived for the most part as the record of the evolution of important mathematical ideas. This is a useful approach, especially for the mathematician. But again, because it remains wholly within the confines of mathematics itself, it lends credence to the myth of isolated mathematics. It is possible to look at the past in terms of what problems or types of problems were of compelling interest. This point of view affords more opportunity to examine the relationship between mathematics and other aspects of culture.

To take an example which may be farfetched, is it mere coincidence that the curbing of the unrestrained manipulation of infinite series coincided with a widespread rejection of Hegel's speculative philosophy? A less fantastic example is the importance for Plato of the problem of duplicating a cube. On this see van der Waerden [**36**], p. 138.

For an example from the more recent past, one must be struck by the fact that the logical foundations of mathematics seemed to crumble at nearly the same time as faith in cultural values and 'progress' was disintegrating.

6. The distinction between pure and applied mathematics and the idea of mathematics as a tool are read into the history of mathematics. This is closely related to the first point mentioned. It has the effect of rendering invisible the problematic character of the distinction (and of 'toolhood').

For example, we often read that mathematics began in ancient Egypt as applied mathematics. Indeed, this is shown by the very name: "geometry" means earth measurement. (Of course this name is Greek, not Egyptian, and reports the Greek understanding of what the Egyptians were doing.) To call an activity "applied mathematics" is to *appeal* to our distinction rather than to show that it is applicable. That distinction is based on the possibility of using mathematics as a tool with which to understand and manipulate nature. But this possibility exists only within the modern Western conception of nature, a conception which appears to be totally foreign to the ancient Egyptians, the ancient Greeks, and even the medieval Christians.

Philosophy. There are many who will claim that in passing from history to philosophy we pass from a light mist into a dense fog. Our historical remarks may have some relevance to the understanding and teaching of mathematics but what can we expect from philosophy? The chief interest in the philosophy of mathematics arose from the 'foundational crisis' provoked by the discovery of antinomies in set theory. This has been straightened out; indeed the effort has produced a robust mathematical logic. Mathematics has emerged from its crisis all the stronger and hardly seems in need of a 'philosophy'. There is something presumptuous about the very idea of a philosophy of mathematics. This prattle about Platonism, realism and the like—is there any point to it? Was Archimedes or Newton a philosopher? or Gauss? Philosophers can talk forever; mathematics *works*.

The differences between philosophy, history and mathematics seem to rest on the relative clarity of their problems. Philosophical problems are never clear; this is why positivists can do without philosophy. Historical problems run through a range from clear but trivial, e.g., who was the thirteenth president of the U.S.?, toward the philosophical, e.g., how has the abundance of natural resources in the U.S. influenced the evolution of our political system? In contrast to problems of this sort a mathematical problem is clear. It may be very difficult but we believe we will know when it's solved. There are proofs which are extremely complex (e.g., the classification of finite groups) but there seems to be a qualitative difference between complexity and the sort of obscurity which attends philosophical problems. Mathematical problems call for solutions; philosophical problems call for thought. An adequate philosophy of mathematics will show that this difference is not as sharp as I have portrayed it.

What is needed is a discussion that begins by examining what commonly passes as philosophy of mathematics. It should sift through the various problems and separate the trivial from the essential. Such a discussion must avoid the methodological naivety of thoughtlessly using the very ideas it examines. It should issue in a more profound understanding of mathematics, particularly of the role (for the most part invisible) that mathematics plays in our most basic conceptions of what is and what it means to question. In the present context this can be done only in the most sketchy form.

The issues of a philosophy of mathematics may be grouped under the following general headings: logical foundations, loss of certainty, the nature of proof, the relation of mathematical knowledge to the 'real world', and the 'ontological status' of mathematical objects such as numbers, sets, functions, etc. In particular the issue of logical foundations does not cover all of the philosophy of mathematics.

Preoccupation with foundational questions has forced the other issues out of view and has led to the trivialization of the philosophy of mathematics which is effective in the attitudes of most working mathematicians. Reuben Hersh (in [**18**]) describes the working mathematician as "a Platonist on weekdays and a formalist on Sundays." (By formalism he means "the philosophical position that much or all of pure mathematics is a meaningless game.") When mathematicians are unable to explain how their work has meaning they resort to calling it meaningless. But this is absurd; it's like suggesting that someone can not really recognize an old friend unless he can describe in a systematic way how he does it.

It seems obvious to me that there is no such thing as a meaningless game; it is equally obvious that pure mathematics is meaningful. As Hersh suggests, the choice between Platonism and formalism is too meager (it remains so when you include "intuitionism" and "logicism") and the mathematician should not be asked to make it. It is a philosophical problem to clarify the *phenomenon* of meaning in mathematics; an important one in my opinion.

Perhaps now that it is no longer *the* pressing issue to provide a logical foundation for mathematics we can ask about the meaning of this demand. What assumptions about the nature of mathematics and its relation to thought and experience in general are implicit in the demand for logical foundations? This is a question which can not be answered by producing a new

definition of the word "set" or even of "foundation". It calls more for thought than for 'an answer'.

As one thinks about this problem there emerges a better understanding of problems in general; an understanding which can be useful to us in the classroom. At the same time that a problem asks for something (the 'unknown') it also specifies—sometimes explicitly but more often implicitly—a framework (which is not put into question) in terms of which or on the basis of which the solution is to be sought. Many problems in algebra and calculus are very simple in this respect. They are comparable to the problem of assembling a jigsaw puzzle when you know what the result should look like. The picture of the final result which appears on the box containing the puzzle is part of the framework within which one approaches the task. (The same is true of the assumption that the pieces do in fact fit together.)

If there is no picture on the box the problem is more difficult but is still possible. More of the framework is put into question if we are simply presented with a box of pieces and asked to assemble them so that they constitute a picture. Here the framework is reduced to the notions of what it means for pieces to fit together and what the word "picture" means.

The question about the foundations of mathematics appears to put into question something called the "nature of mathematics" on the basis of a framework of "thought and experience in general". The vagueness of the terms is an irremovable obstacle; it is an *essential* part of the problem. What is perhaps more important is that the more we think about it the clearer it becomes that we can not put into question the nature of mathematics without also disturbing the ideas of thought and experience in general.

This small insight does not solve the problem; it tends to shift it. It seems to me most fruitful to ask of the philosopher that he *make visible* the role that mathematics plays in our conception of thought and experience. This has a necessarily circular character which I think of as analogous to successive approximation. One begins with a crude idea of mathematics (the science of numbers and their relations for example) and a crude idea of experience (perception of material objects). But already here as soon as we have distinguishable objects a 'mathematical' idea, quantity, is implicit. Thinking about this leads to a little bit clearer idea of mathematics, etc.

The recognition that mathematics participates in experience to a much greater extent than we commonly think is perhaps the most important result of thinking about foundations. It has important ramifications for what I called above the relation of mathematical knowledge to the real world. (The 'real world' is that in which we eat and drink, cut grass, and hold most of our conversations. It is a world of material objects from which we look out to a world of ideas. It is a useful and fascinating metaphysical relic deserving of some veneration. It can also be a mischief maker.) Mathematics seems to grow from within itself without reference to the real world. And yet, again and again there arise applications for what was previously considered 'pure' mathematics. We believe that mathematics exists in two varieties, the pure and the applied. Davis and Hersh ([**4**], p. 86) quote Hardy's 'avowal of purity' and add "[it is a dominant ethos of twentieth-century mathematics] that the highest aspiration in mathematics is the aspiration to achieve a lasting work of art. If, on occasion, a beautiful piece of pure mathematics turns out to be useful, so much the better." Kline's book [**21**] on the other hand amounts to an impassioned argument that the proper roots and nourishment of mathematics lie in nature and applications.

We can arrive at some understanding of the issue of pure versus applied mathematics by first looking more carefully at what the distinction takes for granted. The quote from Davis and Hersh suggests a distinction between aesthetic-based and use-based evaluations. This is an extremely complicated situation from which I wish to draw out only one element: the aesthetic appreciation of an object (such as a proof of a theorem) is usually seen as an appreciation of the object itself; the object which is appreciated for its usefulness derives its value from that of the task for which it is useful. In the idea of applied mathematics there is a likeness to a tool.

This fact has its uses in the classroom. For many students it is very helpful to realize the similarity between how one approaches a mathematical problem and how one approaches a mechanical one. For example a comparison can be made between such methods of solving quadratic equations as trial and error or factoring and such methods of making holes as using a nail and hammer or using a push-drill. The quadratic formula is as useful in its way as an electric drill. Seeing the tool-aspect of mathematics in problem solving sometimes helps remove the aura of mystery which often accompanies the student's perception of mathematics.

But this conception of mathematics as a tool, which is fundamental to our idea of applied mathematics, still needs to be examined. Ordinarily we think of a tool as something inconspicuous which we use when the task demands it and then lay aside. For a time the role of language in articulating thought was likened to that of a tool but it was soon recognized that language plays a much more active role than the tool analogy suggests. In fact, I submit that the ordinary idea of a tool is one of those first-order approximations that function marvelously well in our day-to-day dealings but tend to collapse when asked to carry too much conceptual weight. Even on the level of purely mechanical tasks tools play an active role in the conception of problems. The mechanic's familiarity with what tools are available conditions his perception of a mechanical task just as the chess player's knowledge of how the pieces move conditions his perception of the chessboard.

The active role played by mathematical ideas in the 'perception' of applied problems is so great that I think the tool analogy is entirely misleading in this context. The distinction between pure and applied mathematics not only misses this fact; it covers it over. Certainly we can find 'pure' topics in mathematics and other topics which are used as passively as we ordinarily use a screwdriver. But we must not let that blind us to the very important role that mathematics plays in determining what counts for us as a problem and what would count as a solution. Once this role begins to be recognized the distinction between pure and applied mathematics begins to disintegrate along the common edge.

Wittgenstein puts this very neatly: "But what things are 'facts'? Do you believe that you can show what fact is meant by, e.g., pointing to it with your finger? Does that of itself clarify the part played by 'establishing' a fact?—Suppose it takes mathematics to define the *character* of what you are calling a 'fact'!... Why should not mathematics, instead of 'teaching us facts', create the forms of what we call facts?" (See [**38**], p. 381.)

Loss of certainty. The story of the loss of certainty is told by Morris Kline in his book *Mathematics: The Loss of Certainty*, [**21**]. He describes the emergence of the idea of God's mathematical design of nature during the 16th century as the result of tension between current religious belief and the reawakening interest in classical thought. Understanding of mathematical truth became tantamount to insight into God's design. This was a golden age for mathematics. Its results were as far beyond doubt as the possibility of a split between pure and applied aspects.

The remainder of the book describes the gradual loss of innocence. Mathematics reveals itself increasingly as a human creation and even as subject to human fallibility. It stood for a while as the last bastion of certainty, the final warrant for a belief in the possibility of knowledge, and then fell. From its remains there issued a squabble over foundations which has now subsided without genuine resolution. We mathematicians cannot hope to emerge from the shadow of Gödel's Theorems. "The efforts to eliminate possible contradictions and establish the consistency of the mathematical structures have thus far failed." ([**21**], p. 276).

Ostensibly this loss of certainty presents serious problems of a "philosophical" sort which are quite obviously related to mathematics. If indeed mathematics has served philosophers from Plato to Kant as a model for knowledge in general, then perhaps a change in the conception of mathematics should alter the understanding of that tradition. More relevant to us as mathemati-

cians and teachers is the fact that our own understanding of what mathematics is has been conditioned by that same tradition (Plato to Kant). This results in a tension between our direct contact with mathematics and our understanding of it which is mediated by its recirculation through the tradition. This tension manifests itself in what Davis and Hersh call the Philosophical Plight of the Working Mathematician ([**4**], p. 321).

As far as the loss of certainty is concerned in itself (i.e., not as a historical-cultural phenomenon) it does not seem extraordinarily surprising or significant to me. I am far more puzzled by what 'absolute certainty' might mean than by the fact that mathematics doesn't offer it. It seems to me that there is a similarity between this historical event (of 450 years duration) and the debating tactic which builds and then destroys a straw man.

Mathematics is a human creation. That does not mean that it is arbitrary, but it does mean that it would be immodest to expect it to be 'certainly true' in the common sense of the phrase. It is incoherent to try and imagine mathematics as a source or body of absolutely certain knowledge. To say that a proof demonstrates the truth of a theorem is to say that it warrants the belief that the theorem will not at some future time be a party to a contradiction. The history of mathematics shows that standards of rigor in proof have varied and it affords examples of 'false proofs' which were widely accepted and 'correct proofs' which were widely rejected.

These examples indicate what we may call the "human element" in proof (in addition to the most obvious such, that proofs are made by humans). Proof is a form of mathematical discourse. It functions to unite mathematicians as practitioners of one mathematics. It is sometimes suggested that by formalizing proofs and building proof-checking machines we can reduce the influence of human fallibility on mathematical proofs. No doubt; but if this is seen as the first step of a surreptitious program to resurrect the 'absolute truth' of mathematics, it is doomed to failure. In fact, thinking about formal proofs and proofcheckers gives an interesting perspective on the human aspect.

First of all new, 'trivial', sources of error present themselves: errors in encoding the proof, errors from fatigue in dealing with the sheer complexity of some proofs, and even possible errors in the machine. Secondly and more importantly, a proof functions in mathematics only when it is *accepted* as a proof. This acceptance is a behavior of practicing mathematicians. Even if the verdict of a particular proofchecker comes to be accepted by the community of mathematicians they will reserve the right to reconsider and to admit nonformal proofs.

Fermat wrote that "the essence of proof is that it compels belief." To the extent that the compulsion operates via insight, (relatively) informal proofs will continue to play an important role in mathematics. Proofs that yield insight into the relevant concepts are more interesting and valuable to us as researchers and teachers than proofs that merely demonstrate the correctness of a result. We like a proof that brings out what seems to be essential. If the only available proof of a result is one that seems artificial or contrived it acts as an irritant. We keep looking and thinking. Instead of being able to move on, we are arrested. I mention these familiar facts only to emphasize that proof is not merely a system of links among various theorems, axioms, and definitions but also a system of discourse among people concerned with mathematics. As such it functions in a variety of ways.

Mathematical objects. Traditionally there has been much debate about the existence of mathematical objects such as numbers, categories, etc. This is closely related to the question of whether mathematical results are discovered or invented. It is extremely puzzling how these questions have managed to command as much attention as they have. In the first place it is hard to see what would stand or fall on the basis of a decision. Secondly it seems impossible to entertain the question seriously without making the tacit assumption that there is somehow a ground of 'absolute truth' from which to debate the issue. But there is no more reason to suppose such a ground than to imagine that we would recognize it if it came into view.

In [9] Gödel discussed the objects of set theory: "despite their remoteness from sense experience, we do have something like a perception also of the objects of set theory.... I don't see any reason why we should have less confidence in this kind of perception... than in sense perception, which induces us to build up physical theories and to expect that future sense perceptions will agree with them... ." He continues by pointing out that the objects of physical experience are no more immediately given than are the objects of mathematics. This passage asserts a 'Platonistic' position (especially a little beyond what I have quoted) but I like it for other reasons. First, because it makes quite clear that the objects of physical experience constitute the underlying model for our concept of existence. The more 'refined' models are produced by modification and analogy. Second, because it suggests that the object of physical experience, once we pass to the level of theoretical reflection, is just as problematic as the object of mathematics. We want to talk about the existence of mathematical objects because we want to make sense of the obvious fact that we are communicating when we discuss them. But why should their existence be something above and beyond this (possibility of) communication?

Empiricism. The last 'philosophical' issue I want to discuss is that of the difference between mathematics and the empirical sciences. Any reader who has come this far will know that I am not going to pretend to explain how things 'really stand'. One could do so only by begging the question. Rather I hope to draw into view some features of mathematics and empirical science that will render this distinction problematic.

The distinction appears to be founded on the obvious: empirical science has as its object the material reality which is common to and underlies all of human experience while the objects of mathematics are ideas. (The question of who makes the ideas may be left to the side.) In section two of [25] Lakatos describes this distinction as follows. In the traditional (Euclidean) model of mathematics, truth is injected at the top in the form of indubitable axioms and flows downward via deductive channels to particular theorems. On the other hand, empirical science does not begin with indubitable axioms; rather, what is beyond contest is at the bottom, the reality of hard facts. Falsehood flows upward to the theoretical assumptions. (If a theoretical prediction clashes with the facts the falsehood is retransmitted back to the theory which must then be modified.)

Lakatos maintains that Euclidean programs for mathematics have failed and that this failure has stimulated a tendency to see mathematics as more like the empirical sciences. More importantly a careful look at how mathematicians work and how they have been working for four hundred years makes clear that the Euclidean model is simply not descriptive of mathematics. (At most we can say that it is a model for the exposition of a mature part of mathematics; a model whose merits and liabilities are subject for legitimate debate.) In the last fifty years it has even ceased to be descriptive of how philosophers think about mathematics.

During the same time the notions of "hard facts" and "objective reality" have also become increasingly problematic. This has been provoked in part by reflections on developments in modern physics. But already the psychology of ordinary perception provides abundant evidence to undermine the idea of a reality which is immune to theoretical influence. Thus the conceptions of mathematics and empirical science have each been changing in the direction of the other.

The development of algebraic topology affords an example of a deep similarity between mathematics and empirical science. Eilenberg and Steenrod could only write their book, [5], after a period in which ideas had been tried in a variety of combinations. Which combinations were best was decided by the leading problems but also by a developing internal criterion. The test of the axioms is always: do they capture the essentially relevant features of the guiding problems?

The situation seems to me entirely analogous to what has happened sometimes in the so-called human sciences. It often happens that a particular work captures, and thereby identifies, the essence of a certain chaotic development. (Very often this is a judgment which can be made only retrospectively.) Spengler's *Decline of the West* comes to mind; likewise Heidegger's *Being and*

Time. Such events are like a cold crystal in a rain cloud: they form a point of condensation. If the fog is thick enough the condensation spreads quickly and things soon look very different. The analogy goes further though. As time passes a shift takes place: instead of seeing *Being and Time* as the quintessential product of the intellectual climate of Germany in the twenties we begin to see Germany in the twenties as if through the lens of *Being and Time*.

Likewise, once the axioms proposed for a particular mathematical theory (a theory which has to have existed, perhaps in a disorganized way, *before* the axioms could have been formulated) are accepted (a complex phenomenon with a very definite sociological dimension), they begin to *define* that theory.

The development which I have just described within mathematics (and also, to some extent, within the 'human sciences') is very like what happens in empirical science. If we try to imagine a condition utterly without theory, then it seems as though 'the phenomena' simply present themselves. There is and can be no possible thought of illusion or deception. This condition I would refer to as the state of primordial experience. (I do not assert that such a condition has ever existed except perhaps as a theoretical idea.) Reality is a theoretical construct; it is the hypothetical *x of* which experience *is* experience. The empirical sciences depart from primordial experience; they aim to articulate it as a coherent, intelligible whole. (The word "whole" is important. Eventually, for example, the physical, biological, and ethical models of man must interlock perfectly.) Such ideas as object, material, force, organism, ecosystem, etc., are to the empirical sciences what the axioms are to a particular part of mathematics. They serve to organize experience and to make comparisons (and science) possible. As such they remain accountable to experience although this fact is obscured because they are also the basic elements in terms of which we elaborate that experience.

Puzzlement over the difference between mathematics and empirical science originates to a great extent in the fact that it is much more difficult for us to see the material object as a theoretical construct than, for example, the integer three. It is enhanced by the fact that we use sets of objects to found our formal definition of number and thereby give to objects a more 'fundamental' status. Finally, the fact that we write science in the language of mathematics influences us to think of mathematics as a free creation of the human spirit in contrast to science which is 'anchored to the facts'. I hope that my remarks have weakened this distinction.

Conclusion. In closing I must emphasize that I do not advocate teaching history or philosophy in mathematics courses. Rather, I maintain that the *teacher* needs to have a thorough acquaintance with them in order to teach more effectively. I will be criticized for not being more explicit about the mechanics of how understanding history and philosophy (in addition to mathematics) translates into more effective teaching. In the present context my response must be to appeal to an already familiar situation. When you teach a course which is not packaged in advance you need to know vastly more about the topic than you will ever discuss in detail. Otherwise there is no basis from which to choose which of many interesting beginnings will be most fruitful to pursue. Some have led to dead ends in the literature and some may lead up to the edge of current research. To be effective you need to have already explored these paths, not just the one that ends up being taken. Similarly, in our lower level courses we are teaching students how to solve certain types of problems. For the students the techniques at first seem remote from ordinary experience; for many of them mathematics is indeed a foreign language. We can not take the time to show them in detail the role mathematics has played in creating the world they take for granted and which seems to them so remote from the abstractions they are asked to master. But our awareness of that role can affect our approach at every turn; it manifests itself as an ease and breadth of understanding which is more convincing than speeches.

Acknowledgments. I have benefited from criticism of an earlier version of this essay for which I am grateful to Lloyd Chambless.

References

1. Paul Benacerraf and Hilary Putnam, eds., Philosophy of Mathematics, Prentice-Hall, Englewood Cliffs, 1964.

2. Brand Blanshard, The Nature of Thought, Humanities Press, New York, 1939

3. Georg Cantor, Beiträge zur Begrundung der transfiniten Mengenlehre, Math. Ann., 46 (1895) 481–412. English translation: Contributions to the Founding of the Theory of Transfinite Numbers, Dover, New York, 1955.

4. Philip J. Davis and Reuben Hersh, The Mathematical Experience, Houghton Mifflin, Boston, 1981.

5. Samuel Eilenberg and Norman Stenrod, Foundations of Algebraic Topology, Princeton University Press, Princeton, 1952

6. Howard Eves, An Introduction to the History of Mathematics, 5th ed., Saunders, Philadelphia, 1983.

7. Paul Feyerabend, Against Method, Verso, London, 1978.

8. Hans-George Gadamer, Vom Zirkel des Verstehens, in [**30**], pp. 24–34.

9. Kurt Gödel, What is Cantor's Continuum Problem?, in [**1**], pp. 258–273.

10. Nicolas D. Goodman, Mathematics as an objective science, this Monthly, 86 (1979) 540–551.

11. Greek Mathematics, tr. by Ivor Thomas, Loeb Classical Library, Harvard University Press, Cambridge, 1967.

12. W. K. C. Guthrie, A History of Greek Philosophy, Vol. 1, The Earlier Presocratics and the Pythagoreans; Vol. 2, The Presocratic Tradition from Parmenides to Democritus, Cambridge University Press, 1962, 1965.

13. Thomas Heath, A History of Greek Mathematics, Vol. 1, Dover, New York, 1981.

14. Martin Heidegger, Die Frage nach dem Ding, Niemeyer, Tübingen, 1962. English translation: What is a thing, Henry Regnery, Chicago, 1967.

15. ______, Die Zeit des Weltbilds, in Holzwege, 6th ed., Klosterman, Frankfurt am Main, 1980; English translation: The Age of the World Picture, in The Question Concerning Technology, Harper, New York, 1977.

16. ______, Sein und Zeit, 14th ed., Max Niemeyer, Tübingen, 1977. English Translation: Being and Time, Harper and Row, New York, 1962.

17. Werner Heisenberg, Grundlegende Voraussetzungen in der Physik der Elementarteilchen, in [**30**], pp. 291–297.

18. Reuben Hersh, Some proposals for reviving the philosophy of mathematics, Adv. in Math., 31 (1979) 31–50.

19. Douglas R. Hofstadter, Gödel, Escher, Bach: An Eternal Golden Braid, Basic Books, New York, 1979.

20. Philip Kitcher, The Nature of Mathematical Knowledge, Oxford University Press, New York, 1984.

21. Morris Kline, Mathematics: The Loss of Certainty, Oxford University Press, Oxford, 1980.

22. Stephan Körner, On the relevance of post-Gödelian mathematics to philosophy, in [**27**], pp. 118–132.

23. Edna E. Kramer, The Nature and Growth of Modern Mathematics, Princeton University Press, Princeton, 1981.

24. Thomas S. Kuhn, The Structure of Scientific Revolutions, 2nd ed., University of Chicago Press, Chicago, 1970.

25. Imre Lakatos, A renaissance of empiricism in the recent philosophy of mathematics?, in Mathematics, Science and Epistemology, pp. 24–42.

26. ______, Mathematics, Science and Epistemology (Philosophical Papers, Vol. 2), Cambridge University Press, Cambridge, 1978.

27. ______, (ed.) Problems in the Philosophy of Mathematics (Proceedings of the International Colloquium in the Philosophy of Science, London, 1965) North-Holland, Amsterdam, 1967.

28. ______, Proofs and Refutations, Cambridge University Press, Cambridge, 1976.

29. Saunders Mac Lane, Mathematical Models: A sketch for the philosphy of mathematics, this MONTHLY, 88 (1981) 462–472.

30. Günther Neske, ed., Martin Heidegger zum Siebzigsten Geburtstag, Festschrift Neske, Pfullingen, 1959.

31. George Pólya, Mathematical Discovery, Wiley, New York, 1981.

32. Joachim Ritter, ed., Axiom, in Historisches Wörterbuch der Philosophie, Vol. 1, Schwabe, Basel, 1971, pp. 737–748.

33. Karl J. Smith, The Nature of Mathematics, 4th ed., Brooks/Cole, Monterey, 1984.

34. Oswald Spengler, The Decline of the West, Vol. 1, Knopf, New York, 1926.

35. B. L. van der Waerden, La démonstration dans les sciences exactes de l'antiquité, Bulletin de la Société Mathématique de Belgique, lx (1957) 8–13.

36. ______, Science Awakening, Vol. 1, Noordhoff, Groningen, 1954.

37. André Weil, Number Theory, An Approach through History, Birkauser, Boston, 1983.

38. Ludwig Wittgenstein, Remarks on the Foundations of Mathematics, revised ed., MIT, Cambridge, 1978.

THE WILLIAM LOWELL PUTNAM MATHEMATICAL COMPETITION

LEONARD F. KLOSINSKI
Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053

G. L. ALEXANDERSON
Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053

LOREN C. LARSON
Department of Mathematics, St. Olaf College, Northfield, MN 55057

The following results of the forty-sixth William Lowell Putnam Mathematical Competition, held on December 7, 1985, have been determined in accordance with the governing regulations. This annual contest is supported by the William Lowell Putnam Prize Fund for the Promotion of Scholarship, left by Mrs. Putnam in memory of her husband, and is held under the auspices of the Mathematical Association of America.

The first prize, five thousand dollars, was awarded to the Department of Mathematics of **Harvard University**, Cambridge, Massachusetts. The members of the winning team were: Glenn D. Ellison, Douglas S. Jungreis, and Michael Reid; each was awarded a prize of two hundred fifty dollars.

The second prize, two thousand five hundred dollars, was awarded to the Department of Mathematics of **Princeton University**, Princeton, New Jersey. The members of the winning team were: Michael A. Abramson, Douglas R. Davidson, and James C. Yeh; each was awarded a prize of two hundred dollars.

The third prize, one thousand five hundred dollars, was awarded to the Department of Mathematics of the **University of California**, Berkeley, California. The members of the winning team were: Michael J. McGrath, David P. Moulton, and Jonathan E. Shapiro; each was awarded a prize of one hundred fifty dollars.

The fourth prize, one thousand dollars, was awarded to the Department of Mathematics of **Rice University**, Houston, Texas. The members of its team were: Charles R. Ferenbaugh, Thomas M. Hyer, and Thomas M. Zavist; each was awarded a prize of one hundred dollars.

The fifth prize, five hundred dollars, was awarded to the Department of Mathematics of the **University of Waterloo**, Waterloo, Ontario, Canada. The members of its team were: David W. Ash, Yong Yao Du, and Kenneth W. Shirriff; each was awarded a prize of fifty dollars.

The five highest-ranking individual contestants, in alphabetical order, were **Martin V. Hildebrand**, Williams College; **Everett W. Howe**, California Institute of Technology; **Douglas S. Jungreis**, Harvard University; **Bjorn M. Poonen**, Harvard University; and **Keith A. Ramsay**, University of Chicago. Each of these students was designated a Putnam Fellow by the Mathematical Association of America and awarded a prize of five hundred dollars by the Putnam Prize Fund.

The next five highest-ranking individuals, in alphabetical order, were *David W. Ash*, University of Waterloo; *Waldemar P. Horwat*, Massachusetts Institute of Technology; *Greg J. Kuperberg*, Harvard University; *John M. Steinke*, Rice University; and *David I. Zuckerman*, Harvard University. Each of these students was awarded a prize of two hundred fifty dollars.

The following teams, named in alphabetical order, received honorable mention: *California Institute of Technology*, with team members Leland F. Brown, Art Duval, and Daniel E. Loeb; *Rensselaer Institute of Technology*, with team members Brenden A. Del Favero, John D. Moores, and Bart C. Vashaw; *Rutgers University*, New Brunswick, with team members Scott E. Axelrod, Liaw Huang, and William M. Wells; *Washington University*, St. Louis, with team members Anders W. McCarthy, Daniel N. Ropp, and Dougin A. Walker; and *Yale University*, with team members Thomas Andrews, Ramzi R. Khuri, and David R. Steinsaltz.

Honorable mention was achieved by the following thirty-four individuals, named in alphabetical order: *Michael A. Abramson*, Princeton University; *Jorg Anthony Brown*, Colorado State

University; *Glenn G. Chappell*, University of Kansas; *Stanley Chen*, California Institute of Technology; *John J. Chew*, University of Toronto; *Constantine N. Costes*, Harvard University; *Douglas R. Davidson*, Princeton University; *Glenn D. Ellison*, Harvard University; *Adam F. Falk*, University of North Carolina, Chapel Hill; *Christopher P. Grant*, Brigham Young University; *Mark C. Hamburg*, University of Michigan, Ann Arbor; *Liaw Huang*, Rutgers University, New Brunswick; *Thomas M. Hyer*, Rice University; *William Carl Jockusch*, Carleton College; *Daniel W. Johnson*, Rose-Hulman Institute of Technology; *Jamshid Mahdavi*, Carnegie-Mellon University; *Michael J. McGrath*, University of California, Berkeley; *David J. Moews*, Harvard University; *David P. Moulton*, University of California, Berkeley; *Lee A. Newberg*, Massachusetts Institute of Technology; *Ken E. Newman*, Washington University, St. Louis; *Steve Newman*, University of Michigan, Ann Arbor; *Michael Reid*, Harvard University; *Daniel N. Ropp*, Washington University, St. Louis; *Randall G. Rose*, Princeton University; *David B. Secrest*, University of Illinois, Champaign-Urbana; *Robert E. Shapire*, Brown University; *Randall D. Smith*, University of Chicago; *Eric H. Veach*, University of Waterloo; *Minh Tue Vo*, University of Waterloo; *Dougin A. Walker*, Washington University, St. Louis; *Christopher S. Welty*, University of California, Berkeley; *James C. Yeh*, Princeton University; and *Thomas M. Zavist*, Rice University.

The other individuals who achieved ranks among the top 102, in alphabetical order of their schools, were: Bethel College, *Jonathan P. McCammond*; University of British Columbia, *Marvin S. Lee*; California Institute of Technology, *Eric K. Babson*, *William D. Banks*, *Leland F. Brown*, *Earl A. Hubbell*, *Kenneth F. Kelley*, *James T. Liu*, *David J. Nice*, *Theron W. Stanford*, *Steven B. Waltman*; University of California, Berkeley, *Jonathan E. Shapiro*; University of California, Davis, *John B. Boyland*, *Michael P. Quinn*; University of California, Los Angeles, *Joshua R. Zucker*; University of California, Santa Barbara, *Emerson S. Fang*; Carleton University, *Serge Elnitsky*; Case Western Reserve University, *Patrick T. Headley*; University of Chicago, *Robert P. Stingley*; Concordia University, *Chinh Mai*; Cooper Union, *George Dedaj*; Duke University, *Michelangelo Grigni*; Harvard University, *David N. Esch*, *Jonathan L. Feng*, *Howard M. Pollack*; Université Laval, *Mario Bergeron*; Massachusetts Institute of Technology, *Jonathan W. Aronson*, *David T. Blackston*, *Anthony J. Camire*, *James R. Rauen*; University of Minnesota, Minneapolis, *Bruce W. K. Brandt*; University of New Mexico, *William J. Goldman*; State University of New York, Buffalo, *Eric J. Cockayne*; State University of New York, Stony Brook, *Robert A. Hockberg*; Oberlin College, *Gregory S. Ludwig*; Princeton University, *Stephen J. Fromm*; Queen's University, *Neale Ginsburg*; Rensselaer Institute of Technology, *Brendan A. Del Favero*; Rice University, *Charles R. Ferenbaugh*; University of Rochester, *Sze-Him Ng*; Rutgers University, New Brunswick, *Scott E. Axelrod*, *William M. Wells*; Texas A & M University, *Mark G. Yarbrough*; University of Toronto, *Gary F. Baumbartner*, *William J. Rucklidge*; Vanderbilt University, *Byron Lee Walden*; Washington State University, *Dale A. Nichols*; University of Washington, *Micah E. Fogel*; University of Waterloo, *Yong Yao Du*, *Peter A. Heeman*, *Kenneth W. Shirriff*; Yale University, *Tony J. Fisher*, *Ramzi R. Khuri*, *Kamal F. Khuri-Makdisi*, *Moses G. Klein*, *Richard S. Margolin*, *David R. Steinsaltz*; Youngstown State University, *John P. Dalbec*.

There were 2079 individual contestants from 348 colleges and universities in Canada and the United States in the competition of December 7, 1985. Teams were entered by 264 institutions.

The Questions Committee for the forty-sixth competition consisted of Bruce Reznick (Chair), Richard P. Stanley and Harold M. Stark; they composed the problems listed below and were most prominent among those suggesting solutions.

PROBLEMS

Problem A-1

Determine, with proof, the number of ordered triples (A_1, A_2, A_3) of sets which have the property that

(i) $A_1 \cup A_2 \cup A_3 = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$,

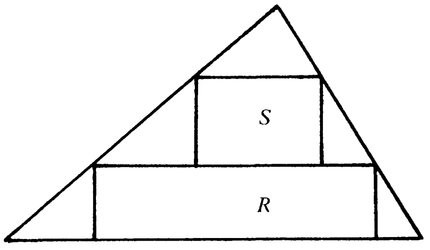
and

(ii) $A_1 \cap A_2 \cap A_3 = \varnothing$,

where $\varnothing$ denotes the empty set. Express the answer in the form $2^a 3^b 5^c 7^d$, where a, b, c and d are nonnegative integers.

Problem A-2

Let T be an acute triangle. Inscribe a pair R, S of rectangles in T as shown:



Let $A(X)$ denote the area of polygon X. Find the maximum value, or show that no maximum exists, of $\dfrac{A(R) + A(S)}{A(T)}$, where T ranges over all triangles and R, S over all rectangles as above.

Problem A-3

Let d be a real number. For each integer $m \geqslant 0$, define a sequence $\{a_m(j)\}$, $j = 0, 1, 2, \ldots$ by the condition

$$a_m(0) = d/2^m, \quad \text{and} \quad a_m(j+1) = (a_m(j))^2 + 2a_m(j), \quad j \geqslant 0.$$

Evaluate $\lim_{n \to \infty} a_n(n)$.

Problem A-4

Define a sequence $\{a_i\}$ by $a_1 = 3$ and $a_{i+1} = 3^{a_i}$ for $i \geqslant 1$. Which integers between 00 and 99 inclusive occur as the last two digits in the decimal expansion of infinitely many a_i?

Problem A-5

Let $I_m = \int_0^{2\pi} \cos(x)\cos(2x) \cdots \cos(mx)\, dx$. For which integers m, $1 \leqslant m \leqslant 10$, is $I_m \neq 0$?

Problem A-6

If $p(x) = a_0 + a_1 x + \cdots + a_m x^m$ is a polynomial with real coefficients a_i, then set

$$\Gamma(p(x)) = a_0^2 + a_1^2 + \cdots + a_m^2.$$

Let $f(x) = 3x^2 + 7x + 2$. Find, with proof, a polynomial $g(x)$ with real coefficients such that

(i) $g(0) = 1$,

and

(ii) $\Gamma(f(x)^n) = \Gamma(g(x)^n)$,

for every integer $n \geqslant 1$.

Problem B-1

Let k be the smallest positive integer with the following property:

There are distinct integers m_1, m_2, m_3, m_4, m_5 such that the polynomial

$$p(x) = (x - m_1)(x - m_2)(x - m_3)(x - m_4)(x - m_5)$$

has exactly k nonzero coefficients.

Find, with proof, a set of integers m_1, m_2, m_3, m_4, m_5 for which this minimum k is achieved.

Problem B-2

Define polynomials $f_n(x)$ for $n \geqslant 0$ by $f_0(x) = 1$, $f_n(0) = 0$ for $n \geqslant 1$, and

$$\frac{d}{dx}(f_{n+1}(x)) = (n+1)f_n(x+1)$$

for $n \geqslant 0$. Find, with proof, the explicit factorization of $f_{100}(1)$ into powers of distinct primes.

Problem B-3

Let

$$\begin{matrix} a_{1,1} & a_{1,2} & a_{1,3} & \cdots \\ a_{2,1} & a_{2,2} & a_{2,3} & \cdots \\ a_{3,1} & a_{3,2} & a_{3,3} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{matrix}$$

be a doubly infinite array of positive integers, and suppose each positive integer appears exactly eight times in the array. Prove that $a_{m,n} > mn$ for some pair of positive integers (m, n).

Problem B-4

Let C be the unit circle $x^2 + y^2 = 1$. A point p is chosen randomly on the circumference C and another point q is chosen randomly from the interior of C (these points are chosen independently and uniformly over their domains). Let R be the rectangle with sides parallel to the x-and y-axes with diagonal pq. What is the probability that no point of R lies outside of C?

Problem B-5

Evaluate $\int_0^\infty t^{-1/2}e^{-1985(t+t^{-1})}\,dt$. You may assume that $\int_{-\infty}^{\infty}e^{-x^2}\,dx = \sqrt{\pi}$.

Problem B-6

Let G be a finite set of real $n \times n$ matrices $\{M_i\}$, $1 \leqslant i \leqslant r$, which form a group under matrix multiplication. Suppose that $\sum_{i=1}^{r}\operatorname{tr}(M_i) = 0$, where $\operatorname{tr}(A)$ denotes the trace of the matrix A. Prove that $\sum_{i=1}^{r} M_i$ is the $n \times n$ zero matrix.

SOLUTIONS

In the 12-tuples $(n_{10}, n_9, \ldots, n_0, n_{-1})$ following each problem number below, n_i for $10 \geqslant i \geqslant 0$ is the number of students among the top 201 contestants achieving i points for the problem and n_{-1} is the number of those not submitting solutions.

A-1. (125, 6, 0, 0, 0, 0, 0, 0, 0, 0, 61, 9)

Every integer, $1 \leqslant i \leqslant 10$, falls into one of six mutually disjoint classes: $A_1 \cap \bar{A}_2 \cap \bar{A}_3$, $\bar{A}_1 \cap A_2 \cap \bar{A}_3$, $\bar{A}_1 \cap \bar{A}_2 \cap A_3$, $\bar{A}_1 \cap A_2 \cap A_3$, $A_1 \cap \bar{A}_2 \cap A_3$, and $A_1 \cap A_2 \cap \bar{A}_3$; hence there are $6^{10} = 2^{10}3^{10}$ different ordered triples.

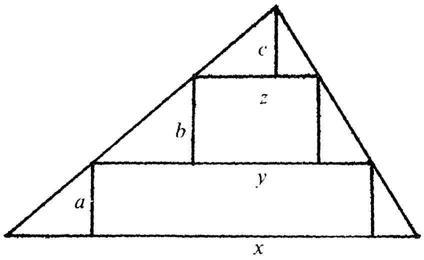
A-2. (29, 15, 31, 14, 38, 11, 2, 15, 6, 4, 21, 15)

Label lengths as in the figure on page 624. Then

$$\frac{A(R) + A(S)}{A(T)} = \frac{ay + bz}{hx/2},$$

where $h = a + b + c$, the altitude of T. By similar triangles,

$$\frac{x}{h} = \frac{y}{b+c} = \frac{z}{c},$$



so

$$\frac{A(R)+A(S)}{A(T)}=\frac{a\dfrac{(b+c)x}{h}+b\dfrac{cx}{h}}{hx/2}=\frac{2}{h^2}(ab+ac+bc).$$

We need to maximize $ab+ac+bc$ subject to $a+b+c=h$. One way to do this is first to fix a, so $b+c=h-a$. Then

$$ab+ac+bc=a(h-a)+bc,$$

and bc is maximized when $b=c$. We now wish to maximize $2ab+b^2$ subject to $a+2b=h$. This is a straight-forward calculus problem giving $a=b=c=h/3$. Hence the maximum ratio is $2/3$ (independent of T).

A-3. (100, 5, 18, 0, 0, 0, 0, 0, 2, 10, 19, 47)

We have $a_n(j+1)+1=(a_n(j)+1)^2$, and hence by induction,

$$a_n(j)+1=(a_n(0)+1)^{2^j}.$$

Therefore

$$\lim_{n\to\infty}a_n(n)=\lim_{n\to\infty}\left(1+\frac{d}{2^n}\right)^{2^n}-1=e^d-1.$$

A-4. (72, 30, 6, 23, 0, 0, 0, 0, 1, 5, 33, 31)

We wish to consider $a_i\equiv 3^{a_{i-1}}\pmod{100}$. Recall that $a^{\phi(n)}\equiv 1\pmod n$ whenever a is relatively prime to n (ϕ is the Euler ϕ-function). Therefore, since $\phi(100)=40$, we can find a_i (mod 100) by knowing a_{i-1} (mod 40). Similarly, we can know a_{i-1} (mod 40) by finding a_{i-2} (mod 16), because $a_{i-1}=3^{a_{i-2}}$ and $\phi(40)=16$. Again, $a_{i-2}=3^{a_{i-3}}$ and $\phi(16)=8$, and therefore

$$a_{i-3}\equiv 3^{a_{i-4}}\equiv 3^{\text{odd integer}}\equiv 3\pmod 8.$$

It follows that $a_{i-2}\equiv 3^3\equiv 11$ (mod 16), and from this,

$$a_{i-1}\equiv 3^{11}\equiv 27\pmod{40},$$

and finally,

$$a_i\equiv 3^{27}\equiv 87\pmod{100}.$$

All of this is valid for all $i\geqslant 4$. Thus, 87 is the only integer that occurs as the last two digits in the decimal expansions of infinitely many a_i.

A-5. (44, 2, 7, 5, 0, 0, 0, 0, 10, 27, 46, 60)

Write

$$I_m = \int_0^{2\pi} \prod_{k=1}^{m} \left(\frac{e^{ikx} + e^{-ikx}}{2} \right) dx = \sum_{\varepsilon_k = \pm 1} \frac{1}{2^m} \int_0^{2\pi} e^{i(\varepsilon_1 + 2\varepsilon_2 + \cdots + m\varepsilon_m)x}\, dx.$$

The integral $\int_0^{2\pi} e^{itx}\, dx$ is zero if t is a nonzero integer and is 2π otherwise. Thus, $I_m \geqslant 0$, and $I_m \neq 0$ if and only if 0 can be written in the form $\varepsilon_1 + 2\varepsilon_2 + \cdots + m\varepsilon_m$ for some $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m \in \{-1, 1\}$. For a sum $\varepsilon_1 + 2\varepsilon_2 + \cdots + m\varepsilon_m$, let r denote the sum of the positive terms and s the sum of the absolute values of the negative terms. Then $r - s = m(m+1)/2$. A necessary condition for $r = s$ is that $m(m+1)/2$ be even; that is, that $m \equiv 0$ or $3 \pmod 4$. Thus, the only candidates satisfying these conditions in $1 \leqslant m \leqslant 10$ are $m = 3, 4, 7$, and 8. We find that $I_m \neq 0$ for each of these because $1 + 2 - 3 = 0$, $1 - 2 - 3 + 4 = 0$, $(1 + 2 - 3) + (4 - 5 - 6 + 7) = 0$, and $(1 - 2 - 3 + 4) + (5 - 6 - 7 + 8) = 0$. (Note that this is easy to generalize to all numbers $m \equiv 0$ or $3 \pmod 4$.)

A-6. (8, 2, 0, 0, 0, 0, 0, 0, 3, 19, 22, 147)

Note that $\Gamma(p(x)) = \int_0^1 |p(e(\theta))|^2\, d\theta$, where $e(\theta) = e^{2\pi i\theta}$. Thus,

$$\Gamma(f(x)^n) = \int_0^1 |f(e(\theta))|^{2n}\, d\theta = \int_0^1 |3e(\theta) + 1|^{2n} |e(\theta) + 2|^{2n}\, d\theta.$$

But

$$|e(\theta) + 2| = |e(\theta)|\,|1 + 2e(-\theta)| = |1 + 2e(-\theta)| = |\overline{1 + 2e(\theta)}| = |1 + 2e(\theta)|.$$

Therefore,

$$\Gamma(f(x)^n) = \int_0^1 |3e(\theta) + 1|^{2n} |2e(\theta) + 1|^{2n}\, d\theta = \Gamma(g(x)^n),$$

where we have set $g(x) = 6x^2 + 5x + 1$.

B-1. (112, 29, 0, 0, 0, 0, 0, 0, 0, 30, 13, 17)

Clearly $k > 1$; otherwise $p(x) = x^5$ and $m_1, \ldots, m_5$ are not distinct. Assume $k = 2$, so $p(x) = x^5 + ax^j$ with $0 \leqslant j \leqslant 4$. We can't have $j \geqslant 2$ since then at least two of the m_i's are equal to 0. Hence $p(x) = x^5 + a$ or $p(x) = x(x^4 + a)$ with $a \neq 0$. But $x^5 + a$ and $x^4 + a$ have at most two real zeros. Therefore $k \geqslant 3$.

Set $m_1 = -2$, $m_2 = -1$, $m_3 = 0$, $m_4 = 1$, $m_5 = 2$. Then

$$p(x) = x(x^2 - 1)(x^2 - 4) = x^5 - 5x^3 + 4x.$$

Hence $k = 3$, and this value of k is achieved for the given m_i's.

B-2. (3, 89, 3, 1, 0, 0, 0, 0, 2, 11, 37, 55)

An examination of low order cases leads one to conjecture that $f_n(x) = x(x+n)^{n-1}$. Clearly this guess satisfies $f_0(x) = 1$, $f_n(0) = 0$ for $n \geqslant 1$. Now

$$\begin{aligned} f'_{n+1}(x) &= (x + n + 1)^n + nx(x + n + 1)^{n-1} \\ &= (n + 1)(x + 1)(x + n + 1)^{n-1} \\ &= (n + 1) f_n(x + 1). \end{aligned}$$

Hence $f_n(x) = x(x+n)^{n-1}$ as guessed. Therefore, $f_{100}(1) = 101^{99}$.

B-3. (95, 15, 15, 11, 0, 0, 0, 0, 5, 3, 23, 34)

Suppose, contrariwise, that $a_{m,n} \leqslant mn$ for all (m, n). Let

$$R(k) = \{(i, j) \colon a_{i,j} \leqslant k\}.$$

By hypothesis, $|R(k)| = 8k$. On the other hand, $ij \leqslant k$ implies that $(i, j) \in R(k)$ and there are

$$\left\lfloor \frac{k}{1} \right\rfloor + \left\lfloor \frac{k}{2} \right\rfloor + \cdots + \left\lfloor \frac{k}{k} \right\rfloor \geqslant \left(\frac{k}{1} - 1 \right) + \cdots + \left(\frac{k}{k} - 1 \right) > k(\log k - 1)$$

such pairs. Hence $8k > k(\log k - 1)$, which is a contradiction for $k > e^9$.

B-4. (115, 30, 9, 1, 2, 6, 4, 0, 9, 8, 11, 6)

Let $p = (\cos\theta, \sin\theta)$ and $q = (x, y)$. The other two vertices of R are $(\cos\theta, y)$ and $(x, \sin\theta)$, so no point of R lies outside of C if and only if $\cos^2\theta + y^2 \leqslant 1$ and $\sin^2\theta + x^2 \leqslant 1$, or equivalently, $|y| \leqslant |\sin\theta|$ and $|x| \leqslant |\cos\theta|$. Note that these conditions imply that (x, y) lies inside the circle, so that, for any θ, the probability that (x, y) satisfies these conditions is

$$\frac{2|\sin\theta|2|\cos\theta|}{\pi} = \frac{2}{\pi}|\sin 2\theta|,$$

and the overall probability is

$$\frac{1}{2\pi}\int_0^{2\pi} \frac{2}{\pi}|\sin 2\theta| d\theta = \frac{1}{2\pi} \cdot \frac{2}{\pi} \cdot 4 = \frac{4}{\pi^2}.$$

B-5. (14, 8, 2, 2, 0, 0, 0, 0, 2, 2, 61, 110)

Let $I(x) = \int_0^\infty t^{-1/2} e^{-at - xt^{-1}}\, dt$, where $a = 1985$. Then

$$I'(x) = -\int_0^\infty t^{-1/2} e^{-at - xt^{-1}} t^{-1}\, dt = -\int_0^\infty t^{-3/2} e^{-at - xt^{-1}}\, dt.$$

Make the substitution $u = 1/t$, and the last equation becomes

$$I'(x) = -\int_0^\infty u^{-1/2} e^{-au^{-1} - xu}\, du.$$

Now let $w = \dfrac{x}{a} u$, and the last equation is

$$I'(x) = -\left(\frac{a}{x}\right)^{1/2} \int_0^\infty w^{-1/2} e^{-xw^{-1} - aw}\, dw = -\left(\frac{a}{x}\right)^{1/2} I(x).$$

Therefore, $\log I(x) = -2(ax)^{1/2} + C$, or equivalently, $I(x) = ke^{-2(ax)^{1/2}}$. Also,

$$k = I(0) = \int_0^\infty t^{-1/2} e^{-at}\, dt = \int_0^\infty 2e^{-at^2}\, dt = \frac{\sqrt{\pi}}{\sqrt{a}}.$$

This yields

$$I(a) = \frac{\sqrt{\pi}}{\sqrt{a}} e^{-2a}.$$

(Note: the integral is essentially the K-Bessel function $K_{1/2}(3970)$.)

B-6. (5, 0, 0, 0, 0, 4, 0, 0, 9, 7, 26, 150)

Let $S = \sum_{i=1}^r M_i$. For any j, $1 \leqslant j \leqslant r$,

$$M_j S = \sum_{i=1}^r M_j M_i = \sum_{i=1}^r M_i = S, \quad \text{and hence} \quad S^2 = \sum_{j=1}^r M_j S = rS.$$

Therefore the minimal polynomial $p(x)$ for S divides $x^2 - rx$ and every eigenvalue of S is either 0 or r. Since $\operatorname{tr}(S) = 0$, every eigenvalue of S is zero. Every eigenvalue of $S - rI$ is $-r$, and therefore $S - rI$ is invertible. Hence, from $S(S - rI) = 0$, we get $S = 0$.

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

A QUINTET OF QUITE QUICKLY QUOTED QUERIES

RICHARD K. GUY

Often the MONTHLY gets problems which are too unsolved to go into the Elementary or Advanced sections, even with stars on. But they may be too brief to stand alone in this section. It would be a pity if they were lost, so some are collected here, even though they have no unifying theme. One connects logic with applied mathematics, one is in elementary number theory and the other three concern various aspects of geometry.

ARE MAXWELL'S EQUATIONS LOGICALLY CONSISTENT?

David Dowe, Department of Mathematics, Monash University, Clayton, Victoria, Australia, 3168 asks, more generally: if we regard hypothesized physical "laws" as mathematical "axioms", can certain physical theories be shown to be logically consistent?

A referee notes that applied mathematicians tend to use a very weak notion of consistency and so would say that Maxwell's equations are trivially consistent because they do have solutions, e.g., plane waves. But the query is presumably made in a strong sense: can one deduce, from Maxwell's equations, a result and also its negation?

WHAT ARE THE SMALLEST ARITHMETIC PROGRESSIONS OF COMPOSITE NUMBERS?

Victor Pambuccian, Str. Viitorului 26, 70266 Bucuresti 9, Romania asks for $a(n)$, the smallest integer a for which there is an integer b, $0 < b < a$ and coprime with a, $(a, b) = 1$, such that

$$a + b, \quad 2a + b, \quad \ldots, \quad na + b$$

are all composite. Here are the first few values of $a(n)$ and the corresponding values of b:

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
$a(n)$	3	5	7	11	11	13	19	19	19	31	31	31	31	31	31	31	31	47	47	61	61	61	61
b	1	4	1	5	5	12	1	1	1	19	19	19	19	19	19	19	19	27	27	47	47	47	47

Pambuccian makes three conjectures:

¿ $a(n)$ is always prime ?

[A near miss for $n = 9$ was $a = 25$, $b = 19$. Your editor believes this conjecture, but Paul Erdős doesn't. A computer might settle our differences.]

¿ For each m, there's an n such that $a(n) = a(n+1) = \cdots = a(n+m)$?

¿ The set of distinct $a(n)$ has zero density in the set of primes ?

[The third conjecture assumes that the first is true.]

HOW BIG A SIMPLEX DO THE EXCENTRES OF A SIMPLEX FORM?

Kazuyuki Hatada, Department of Mathematics, Faculty of Education, Gifu University, Yanagido, Gifu City, Gifu 501-11, Japan lets I_j be the excentre of the n-dimensional simplex $P_0P_1 \dots P_n$ which is opposite to the vertex P_j. [There are $n+2$ hyperspheres which touch all $n+1$ hyperfaces of an n-dimensional simplex. The incentre of the simplex is the centre of the hypersphere inside the simplex. The excentres are the centres of the other $n+1$ hyperspheres. For example, a triangle has an incentre and 3 excentres; a tetrahedron 4 excentres.] He then lets $f(n)$ be the minimum ratio of the contents of the two simplexes $I_0I_1 \dots I_n$ and $P_0P_1 \dots P_n$, taken over all shapes of simplex $P_0P_1 \dots P_n$.

Is $f(n) = 2^n/(n-1)^n$ for $n \geqslant 2$?

Is the minimum attained only when the simplex is regular?

Example: $f(2) = 4$, and this ratio of areas is attained only when the triangle $P_0P_1P_2$ is equilateral.

HOW LIKELY ARE RANDOM POINTS IN THE SQUARE TO BE FAR APART?

H. W. Corley, Department of Industrial Engineering, The University of Texas at Arlington, Arlington, TX 76019 lets $(X_1, Y_1), \dots, (X_n, Y_n)$ be a two-dimensional random sample of size n, where X_i, Y_i, $0 \leqslant i \leqslant n$, are independent uniformly distributed random variables on $(0,1)$.

What is the probability $P\{\min_{i \neq j}[(X_i - X_j)^2 + (Y_i - Y_j)^2] > t\}$?

DO UNIQUE CLOSEST ELEMENTS IMPLY CONVEXITY?

Lawrence R. Weill, Department of Mathematics, California State University, Fullerton, CA 92634 considers a subset S of Hilbert space, H, with the property that, to each element of H there is a unique closest element of S, i.e., for each $x \in H$ there is a unique $\hat{x} \in S$, such that for all $y \in S$, $\|x - \hat{x}\| \leqslant \|x - y\|$, and asks

Is S necessarily convex?

He notes that a converse theorem is well known [e.g., **2**]. If K is a non-empty convex closed subset of H, then for each $x \in H$ there is a unique $\hat{x} \in K$, such that for all $\hat{y} \in K$, $\|x - \hat{x}\| \leqslant \|x - y\|$. The problem is mentioned on p. 95 of [**6**]. It has been solved affirmatively in $\mathbb{R}^n$ by Motzkin [**3**]. Other related papers are [**1**], [**4**], [**5**], [**7**].

References

1. V. L. Klee, A characterization of convex sets, this MONTHLY, 56(1949) 247-249; MR 10, 621.

2. David G. Luenberger, Optimization by Vector Space Methods, p. 69, Wiley, 1969.

3. T. S. Motzkin, Sur quelques propriétés caractéristiques des ensembles convexes, Atti Accad. Naz. Lincei Rend. VI, 21(1935) 562-567; Zbl. 11, 411e.

4. R. R. Phelps, Convex sets and nearest points, Proc. Amer. Math. Soc., 8(1957) 790-797; MR 19, 432a.

5. R. R. Phelps, Convex sets and nearest points II, Proc. Amer. Math. Soc., 9(1958) 867-873; MR 21 #2900.

6. Josef Stoer and Christoph Witzgall, Convexity and Optimization in Finite Dimensions I, Grundlehren der mathematischen Wissenschaften, 163 (1970) Springer-Verlag, N.Y.-Berlin; MR 44 #3707.

7. G. Vergnères, Sur l'unicité du "minimum" de la distance d'un point à un ensemble, Atti Accad. Naz. Lincei Rend. VI, 18(1933) 542-546; Zbl. 8, 339a.

NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.
For instructions about submitting Notes for publication in this department see the inside front cover.

A SIMPLE PROOF OF THE FUNDAMENTAL THEOREM OF FINITE MARKOV CHAINS

GEORGE MALTESE
Mathematisches Institut der Universität Münster, Münster, West Germany

The theorem mentioned in the title is as follows:

THEOREM (Markov). *Suppose $p_{ij} \geqslant 0$ for $i, j = 1,2,\dots,n$ and suppose*

$$\sum_{i=1}^{n} p_{ij} = 1 \quad \text{for } j = 1,2,\dots,n.$$

Then there exist numbers $x_j \geqslant 0$ for $j = 1,2,\dots,n$ satisfying:

$$\sum_{j=1}^{n} x_j = 1 \quad \text{and} \quad \sum_{j=1}^{n} p_{ij} x_j = x_i \quad \text{for } i = 1,2,\dots,n.$$

Recently in his extremely interesting and stimulating paper [3], Joel Franklin gave a pretty proof of the Markov result using an old (1902) theorem of Julius Farkas, which itself is of fundamental theoretical significance in the theory of linear (and nonlinear) programming. Franklin [3, p. 239] mentions that a proof of a special case of Markov's theorem occupies five pages in W. Feller's classic [2] and further comments that the general version above of Markov's theorem is usually proved by using the Perron-Frobenius maximum principle for positive matrices or by using the Brouwer fixed point theorem.

In this brief note we shall show that a powerful fixed point theorem of Markov-Kakutani, whose proof in $\mathbb{R}^n$ is much simpler than any proof of the Brouwer fixed point theorem, has the Markov theorem as an immediate consequence. In order to make this paper self-contained, we shall first present a short proof of the fixed point theorem in question.

Let K be a convex set in $\mathbb{R}^n$. An *affine* mapping T: $K \to K$ is a mapping with the property

$$T(tx + (1-t)y) = tTx + (1-t)Ty$$

for all $x, y \in K$ and $0 \leqslant t \leqslant 1$. For the following fixed point theorem, we shall consider $\mathbb{R}^n$ normed by

$$\|x\| = \max_{1 \leqslant j \leqslant n} |x_j|, \quad \text{where } x = (x_1, \dots, x_n).$$

THEOREM (Markov-Kakutani). *Let T be a continuous affine mapping of the non-empty compact convex set K into itself. Then T has a fixed point.*

Proof. Choose $z \in K$ and let $x_N = \frac{1}{N} \sum_{k=0}^{N-1} T^k z$. Then $x_N \in K$ (by convexity) and since K is compact, there exist $x \in K$ and a convergent subsequence x_{N_p} such that

$$\|x_{N_p} - x\| \to 0 \quad \text{as } p \to \infty.$$

This x is a fixed point. To prove this, note that by compactness $\sup\{\|y\|\colon y \in K\} < \infty$. Therefore

$$\|x_{N_p} - Tx_{N_p}\| = \frac{1}{N_p}\|z - T^{N_p} z\| \leqslant \frac{2}{N_p} \sup\{\|y\|\colon y \in K\} \to 0 \quad \text{as } p \to \infty,$$

so that $\|x - Tx\| = \lim_{p\to\infty}\|x_{N_p} - Tx_{N_p}\| = 0$.

We turn now to the proof of Markov's theorem. Define a convex compact subset K of $\mathbb{R}^n$ as follows:

$$K = \left\{ x = (x_1, \dots, x_n) \in \mathbb{R}^n \colon x_i \geqslant 0 \quad \text{for all } i, \sum_{i=1}^n x_i = 1 \right\},$$

and define a linear mapping $T\colon \mathbb{R}^n \to \mathbb{R}^n$ as follows:

$$Tx = \left(\sum_{j=1}^n p_{1j}x_j, \sum_{j=1}^n p_{2j}x_j, \dots, \sum_{j=1}^n p_{nj}x_j \right).$$

The mapping T is continuous because

$$\|Tx\| = \max_{1\leqslant i\leqslant n} \left| \sum_{j=1}^n p_{ij}x_j \right| \leqslant \|x\| \max_{1\leqslant i\leqslant n} \sum_{j=1}^n p_{ij},$$

so that letting $c = \max_{1\leqslant i\leqslant n} \sum_{j=1}^n p_{ij}$ we have

$$\|Tx\| \leqslant c\|x\|.$$

Claim: *The restriction of T to K is an affine continuous mapping from K to K.*

Since T is linear continuous on $\mathbb{R}^n$, its restriction to K is affine continuous. We need only prove that $x \in K$ implies that $Tx \in K$. Let $x \in K$, $x = (x_1, \dots, x_n)$, then by definition:

$$Tx = \left(\sum_{j=1}^n p_{1j}x_j, \dots, \sum_{j=1}^n p_{nj}x_j \right).$$

For each i we have $\sum_{j=1}^n p_{ij}x_j \geqslant 0$. Furthermore

$$\sum_{i=1}^n \sum_{j=1}^n p_{ij}x_j = \sum_{j=1}^n x_j \sum_{i=1}^n p_{ij} = \sum_{j=1}^n x_j = 1$$

and so $Tx \in K$, proving our claim.

Applying the Markov-Kakutani Theorem we obtain the existence of some $\tilde{x} = (\tilde{x}_1, \dots, \tilde{x}_n) \in K$ with $T\tilde{x} = \tilde{x}$. This means

$$\sum_{j=1}^n \tilde{x}_j = 1 \quad \text{and} \quad \sum_{j=1}^n p_{ij}\tilde{x}_j = \tilde{x}_i \quad \text{for } i = 1, 2, \dots, n,$$

which proves Markov's Theorem.

References

1. N. Dunford and J. Schwartz, Linear Operators, Part I, Interscience Publishers, New York, 1958.
2. W. Feller, An Introduction to Probability Theory and Its Applications, Vol. 1, 3rd ed., Wiley, New York, 1968.
3. J. Franklin, Mathematical methods of economics, this Monthly, 90 (1983) 229–244.

ANSWER TO PHOTO ON PAGE 608

Felix Klein. The picture was taken when he was 29 years old.

A NEW FORMULA FOR π

YURI V. MATIYASEVICH
LOMI, 27, Fontanka, Leningrad, 191011, USSR

RICHARD K. GUY
Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4

The first author [see **4** for example] stated the formula

$$\boxed{\pi = \lim_{m\to\infty} \sqrt{\frac{6\log \operatorname{fcm}(u_1,\dots,u_m)}{\log \operatorname{lcm}(u_1,\dots,u_m)}}} \tag{1}$$

in which, to achieve a balance between numerator and denominator, fcm is used to denote the **formal common multiple**, $u_1u_2 \cdots u_m$ (simply the product) in contrast to the more familiar lcm, or **least common multiple**. The base of the logarithms is immaterial and u_i are **Fibonacci numbers**, defined by $u_0 = 0$, $u_1 = 1$ and $u_{i+1} = u_i + u_{i-1}$.

The Fibonacci numbers crop up in many parts of mathematics, and so also, of course, does π, but perhaps they have never before been so directly related. It is true that

$$\pi = 4\sum_{n=1}^{\infty} \arctan(1/u_{2n+1}) \tag{2}$$

but as soon as it is realized that

$$\arctan(1/u_{2n+1}) = \arctan(1/u_{2n}) - \arctan(1/u_{2n+2})$$

then formula (2) is only a sophisticated way of saying that $\pi/4 = \arctan(1/u_2)$.

In fact the Fibonacci numbers in (1) may be replaced by the members of many other second order recurring sequences. We will prove (1) and see later that the proof extends to what we will call **Chebyshev's form of the prime number theorem for recurring sequences**.

Where did formula (1) come from? The function $\pi(x)$, the number of primes less than or equal to the positive real number x, has been, and still is, one of the most actively studied in mathematics. It is a step function with a jump of 1 at each prime. The prime number theorem states that

$$\pi(x) \sim x/\ln x \tag{3}$$

where $\sim$ means that the ratio of the two sides tends to 1 as $x \to \infty$.

The famous 19th century Russian mathematician, P. L. Chebyshev, introduced two other functions, $\theta(x)$ and $\psi(x)$, which are closely related to $\pi(x)$. The first is

$$\theta(x) = \sum \ln p$$

where the sum is taken over the primes $p \le x$. For example $\theta(1) = 0$, $\theta(2) = \ln 2$, $\theta(3) = \theta(4) = \ln 6$, $\theta(5) = \theta(6) = \ln 30$. This is also a step function, but the steps now grow like $\ln x$. So we might expect that

$$\theta(x) \sim \pi(x)\ln x \tag{4}$$

and this is indeed true. See, for example, Theorem 4.4 in [**1**].

Chebyshev's other function can be defined by

$$\psi(x) = \theta(x) + \theta(x^{1/2}) + \theta(x^{1/3}) + \cdots \tag{5}$$

where the series naturally truncates itself after $\log_2 x$ (to base 2) terms. You can deduce from (3), (4) and (5) that

$$\psi(x) \sim x \tag{6}$$

and this is equivalent to (3) (see Theorem 4.4 in [**1**] again). We will call formula (6) **Chebyshev's**

form of the prime number theorem, even though it was not finally proved until two years after his death.

Now, for any increasing sequence of positive integers, $a_1, a_2, \ldots,$ we can define two functions: $\pi_a(x)$, the number of subscripts $i \le x$ for which a_i is prime, and

$$\psi_a(x) = \ln \operatorname{lcm}(a_1, a_2, \ldots, a_{\lfloor x \rfloor}). \tag{7}$$

The classical case is $a_i = i$, where study of the growth of $\pi_a(x) = \pi(x)$ is essentially equivalent to the study of $\psi_a(x)$, which, in this case, can be seen to be the same as the $\psi(x)$ defined in (5). On the other hand, we can't expect such a close relationship between $\pi_a(x)$ and $\psi_a(x)$ for more general sequences.

For example, for the Fibonacci sequence, it is an old and apparently hard question whether $\pi_u(x) \to \infty$ or not. On the other hand, it is easy to prove that, as $x \to \infty$

$$\ln \operatorname{fcm}(u_1, u_2, \ldots, u_{\lfloor x \rfloor}) \sim \frac{1}{2}x^2 \ln \tau \tag{8}$$

where $\tau = (1 + \sqrt{5})/2$ is the **golden ratio**. Now (7) with $a_i = u_i$ is

$$\psi_u(x) = \ln \operatorname{lcm}(u_1, u_2, \ldots, u_{\lfloor x \rfloor})$$

and if we combine this with (8) we see that formula (1) is equivalent to

$$\psi_u(x) \sim (3x^2 \ln \tau)/\pi^2. \tag{9}$$

It is natural to call this **Chebyshev's form of the prime number theorem for Fibonacci numbers**. In order to prove (9), and hence formula (1), we introduce some notation and recall some well-known properties of Fibonacci numbers.

If p is a prime, denote by $\operatorname{ind}_p n$ the greatest exponent e for which p^e divides n. It is known (see **[5]** for example) that every prime p appears as a factor of some Fibonacci number: let $r(p)$ be the **rank of apparition** of p, i.e., the least subscript i for which p divides u_i. For example $r(2) = 3$, $r(3) = 4$, $r(5) = 5$, $r(7) = 8$, $r(11) = 10$ and $r(13) = 7$. It is also known **[5]** that

$$p \text{ divides } u_i \text{ if and only if } r(p) \text{ divides } i. \tag{10}$$

For example, $r(13) = 7$, and 13 divides $u_7 = 13$, $u_{14} = 377$, $u_{21} = 10946, \ldots,$ all u_i such that 7 divides i, and 13 divides no other u_i. When (10) holds, i.e., when $r(p)$ divides i, it is also true that

$$\operatorname{ind}_p u_{qi} \ge \operatorname{ind}_p u_i \tag{11}$$

with equality if q is not a multiple of p. For example, $r(2) = 3$ and $\operatorname{ind}_2 u_3 = 1$, $\operatorname{ind}_2 u_6 = 3$, $\operatorname{ind}_2 u_9 = 1$, $\operatorname{ind}_2 u_{12} = 4$, $\operatorname{ind}_2 u_{15} = 1$, $\operatorname{ind}_2 u_{18} = 3$.

For the rest of the argument each prime p is considered separately and we write ind n for $\operatorname{ind}_p n$ and r for $r(p)$. From (10) and (11) it is easy to deduce that, for $i \le p^e r$,

$$\operatorname{ind} u_i \le \operatorname{ind} u_{p^e r}. \tag{12}$$

Let $w_0 = 1$, $w_m = \operatorname{lcm}(u_1, \ldots, u_m)$ for $m \ge 1$, and write $v_m = w_m / w_{m-1}$ so that

$$w_m = v_1 v_2 \cdots v_m = \operatorname{fcm}(v_1, \ldots, v_m). \tag{13}$$

Then

$$v_i = \frac{w_i}{w_{i-1}} = \frac{\operatorname{lcm}(w_{i-1}, u_i)}{w_{i-1}} = \frac{u_i}{\gcd(w_{i-1}, u_i)}. \tag{14}$$

Consider the sequence

$$\operatorname{ind} w_1, \quad \operatorname{ind} w_2, \quad \ldots, \quad \operatorname{ind} w_i, \ldots.$$

It has its first jump at the rank of apparition $i = r$, and further jumps only at $i = p^e r$,

$e = 1, 2, \ldots$. Thus,

$$\operatorname{ind} v_i = \begin{cases} \operatorname{ind} w_r & \text{if } i = r \\ \operatorname{ind} w_{p^e r} - \operatorname{ind} w_{p^{e-1}}, & \text{if } i = p^e r,\ e = 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases}$$

Now we are ready to establish the important identity

$$u_i = \prod v_d \tag{15}$$

where the product is over the divisors, d, of i. To this end it is enough to show that, for each prime p,

$$\operatorname{ind} u_i = \sum \operatorname{ind} v_d \tag{16}$$

where the sum is over the divisors, d, of i.

Case 1. $\operatorname{ind} u_i = 0$. By (10), the rank of apparition, r, does not divide i. So if d divides i, r does not divide d, and (10) implies that p does not divide u_d for any d which divides i. So (14) implies that p does not divide v_d. That is, $\operatorname{ind} v_d = 0$ for each d which divides i. This proves (16) in Case 1.

Case 2. $\operatorname{ind} u_i > 0$. By (10), $i = qp^e r$ for some $e \geq 0$, where p does not divide q. By (11),

$$\operatorname{ind} u_i = \operatorname{ind} u_{p^e r} \geq \operatorname{ind} u_k$$

for all $k \leq p^e r$, by (12). So

$$\begin{aligned} \operatorname{ind} u_i &= \operatorname{ind} w_{p^e r} \\ &= \operatorname{ind} w_r + \sum_{j=1}^{e} \{\operatorname{ind} w_{p^j r} - \operatorname{ind} w_{p^{j-1} r}\} \\ &= \operatorname{ind} v_r + \sum_{j=1}^{e} \operatorname{ind} v_{p^j r} = \sum \operatorname{ind} v_d \end{aligned}$$

where the sum is over the divisors d, of i. So (16) is proved in Case 2 and (15) follows in either case.

We can write (15) in the form $\sum \ln v_d = \ln u_i$ and use the Möbius inversion formula (e.g., Theorem 2.9 in [1]) to give

$$\ln v_d = \sum \mu(d/i) \ln u_i. \tag{17}$$

In (17) the sum is over the divisors, i, of d, and $\mu(n)$ is the **Möbius function**, defined by $\mu(1) = 1$, $\mu(n) = 0$ if n contains a repeated prime factor, and $\mu(n) = (-1)^{\omega}$ when n is the product of ω distinct primes.

If we substitute (17) in the logarithm of (13), we get

$$\ln w_m = \sum_{d=1}^{m} \ln v_d = \sum_{d=1}^{m} \sum_{i|d} \mu(d/i) \ln u_i. \tag{18}$$

Binet's formula (see [5] for example)

$$u_i = \{\tau^i - (-\tau)^{-i}\}/(\tau + \tau^{-1})$$

gives

$$\ln u_i = i \ln \tau + \varepsilon_i \tag{19}$$

where

$$|\varepsilon_i| = |\ln\{(1 - (-1)^i \tau^{-2i})/(\tau + \tau^{-1})\}| < 1 \quad (i \geq 1). \tag{20}$$

Substitution of (19) in (18) gives

$$\ln w_m = A + B \tag{21}$$

where

$$A = \sum_{d=1}^{m} \sum_{i|d} \mu(d/i)\, i \ln \tau \tag{22}$$

and

$$B = \sum_{d=1}^{m} \sum_{i|d} \varepsilon_i \mu(d/i). \tag{23}$$

From (20), the inner sum in (23) is less than the number of divisors of d. This number is $2a+1$ or $2a$, according as d is square or not, and a of them are $< d^{1/2}$. In either case the number of divisors is $< 2d^{1/2}$ and

$$B < 2m^{3/2} \tag{24}$$

which we will see is small compared with A when m is large.

In fact, if we take out the constant factor $\ln \tau$ from (22), the inner sum becomes

$$\sum_{i|d} i\mu(d/i)$$

which (see Theorem 2.3 of [**1**]) is Euler's totient function, $\varphi(d)$, the number of integers not exceeding d and prime to it, so that (22) becomes

$$A = \ln \tau \sum_{d=1}^{m} \varphi(d). \tag{25}$$

Mertens proved (Theorem 3.7 in [**1**]) that

$$\sum_{d=1}^{m} \varphi(d) \sim 3m^2/\pi^2 \tag{26}$$

and we now know that

$$\begin{aligned}
\varphi_u(m) &= \ln \operatorname{lcm}(u_1, \dots, u_m) && \text{by definition (7)}\\
&= \ln w_m && \text{by the definition of } w_m\\
&= A + B && \text{by (21)}\\
&\sim A \sim (3m^2 \ln \tau)/\pi^2 && \text{by (24), (25) and (26).}
\end{aligned}$$

We have proved (9), Chebyshev's form of the prime number theorem for Fibonacci numbers. We saw that this is equivalent to our main formula (1).

What other sequences can replace the Fibonacci numbers in formula (1)? We do not know the complete answer, but many other second order recurring sequences will do. The properties that we used were that every prime number divides infinitely many Fibonacci numbers, and that such divisibility was controlled by statements (10) and (11). Similar results hold for the sequences defined by $y_0 = 0$, $y_1 = 1$ and

$$y_{n+1} = cy_n \pm y_{n-1}. \tag{27}$$

The case with c even and sign minus was discussed by Martin Davis [**2**]. Every prime p is a factor of infinitely many y_t, since $y_0 = 0$ and the sequence is periodic, modulo p. [Even if we used different initial values, so that there was a finite set of primes which did not divide any of the y_t, we could ignore such primes since they would not affect the asymptotic results.]

Let r be the rank of apparition of p in the sequence (27) and suppose that p divides y_m where $m = kr + l$, $0 \le l < r$. Then, by lemmas 2.5 to 2.8 of [**2**], p divides y_l and $l = 0$, giving the analog of statement (10) above. The inequality corresponding to (11) follows from Davis's lemmas 2.10 and 2.7, and the analog of Binet's formula from the definition of y_n on page 239 of [**2**]. The sequence (27) with sign plus was considered in [**3**]: matrices were used where Davis used the Pell equation, and the need for c to be even was eliminated.

References

1. T. M. Apostol, Introduction to Analytic Number Theory, Springer-Verlag, 1976; MR 55#7892.
2. Martin Davis, Hilbert's tenth problem is unsolvable, this MONTHLY, 80 (1973) 233–269; MR 56#6465.
3. Yuri V. Matiyasevich, Diophantine Sets, Uspehi Mat. Nauk., 27 (1972) #5(167) 185–222, transl. Russian Math. Surveys, 27 (1972) #5 124–164 MR 56#109.
4. Yuri V. Matiyasevich, Abstracts of the Number Theory Conference, Tbilisi, Sept. 1985.
5. N. N. Vorob'ov, Fibonacci Numbers, Moscow, 1963; English translation of second edition, Heath, Boston, 1963; MR 29#5777.

A CHARACTERIZATION OF DIMENSION

CHARLES S. ALLEN
Department of Mathematics, Drury College, Springfield, MO 65802

The Infinite Log Jam Problem: Suppose that a log jam is formed from a sequence of logs $L_1, L_2, L_3, \ldots$ with the length of L_n being n. A rubber tarpaulin is stretched completely around the log jam. Is it possible for the enclosure formed by the tarpaulin to contain a log of infinite length?

The solution of the infinite log jam problem characterizes a geometric difference between finite and infinite dimensional complete normed linear spaces. The mathematical question proposed is: In a Banach space, does the convex hull of every unbounded set contain a ray or, equivalently, does every unbounded convex set contain a ray? The answer is yes if the space is finite dimensional and no if it is infinite dimensional. The proof is given by the following two lemmas and theorem:

LEMMA 1. *In R^n, the closed convex hull of every unbounded set contains a ray.*

Proof. Let A be an unbounded set in R^n. We may assume without loss of generality that A contains the origin. Choose $a_k \in A$ such that $\|a_k\| > k$ for $k \in N$ and let $b_k = a_k/\|a_k\|$. Each b_k is on the unit sphere and hence the set $B = \cup\{b_k\}$ has an accumulation point x.

Let $\lambda > 0$ and S be an open ball that contains λx. There exists a k sufficiently large so that the line segment joining a_k and the origin intersects S. Thus, λx is an accumulation point of the convex hull of A, and $T = \{\lambda x : \lambda > 0\}$ is the desired ray.

LEMMA 2. *In R^n, if C is convex and $\overline{C}$ contains a ray, then C contains a ray.*

Proof. In the case when $n = 2$, the proof is easy. Now suppose $C \subseteq R^3$ with C convex and $\overline{C}$ containing a ray T. If $C \subseteq T$, then clearly C contains a ray. If C is not a subset of T, then choose $x_1 \in C - T$ and let M_1 be the plane generated by x_1 and T. Again we are confronted with two possibilities. If $C \subseteq M_1$, then we have reduced the problem to the case of R^2. If $C - M_1 \neq \phi$, then choose $x_2 \in C - M_1$. Since $\{x_2\}$ is a compact convex set and M_1 is a closed convex set, there exists a hyperplane P that strictly separates x_2 and M_1. On T, choose a sequence $\{y_k\}$ such that the sequence $\{\|y_k\|\}$ diverges. For each k there exists a point $z_k \in C$ such that $\|z_k - y_k\| < 1$ and z_k is on the same side of P as M_1. Let p_k be the intersection of the line segment $z_k x_2$ and P. The sequence $\{\|p_k\|\}$ diverges; thus $B = \cup\{p_k\}$ is an unbounded set with $B \subseteq P$. The hyperplane P may be considered as a copy of R^2 and hence, by Lemma 1, the closed convex hull of B contains a ray. Since Lemma 2 is valid for R^2, the convex hull of B contains a ray. We obtain the desired result by the convexity of C.

The above techniques extend to higher dimensions and the proof follows by induction.

THEOREM. *A Banach space is finite dimensional if and only if every unbounded convex set contains a ray.*

Proof. Let $B = \{x_k : k \in N\}$ be a set of linearly independent unit vectors in an infinite dimensional Banach space. The convex hull C of the set $\{kx_k : k \in N\}$ is unbounded. Moreover,

if $y \in C$, then the component of x_k in the expansion of y with respect to B is bounded in the interval $[0, k]$. This makes it impossible for C to contain a ray since any ray in the subspace generated by B has the form

$$\left\{ \sum_1^N (a_k + \lambda b_k) x_k : \lambda \geqslant 0 \right\},$$

where N is chosen sufficiently large and at least one of the b_k's is nonzero.

The necessity follows from the two lemmas.

THE JIMMY'S BOOK

STEVEN E. LANDSBURG
Department of Mathematics, Cornell University, Ithaca, NY 14853

In 1974, I had the great good fortune to come to the University of Chicago as a graduate student. Our class soon distinguished itself by a prodigious capacity for bridge and by the offbeat problems that we invented to challenge each other. To the best of my knowledge, none of these problems was ever solved, although a few enjoyed the status of mass obsessions for several years.

None of us thought to record any of the problems or our partial solutions at the time; if we had, they would have filled a small book. By analogy with [**1**], and in honor of a venerable Hyde Park institution, it would have to have been called "The Jimmy's Book". My purpose here is to record a few of the most memorable entries from that book that was never written.

1. Zero-One Matrices. This problem was first posed by Michael Stob, who had heard of its appearance on an exam for high school students! Consider the set of $n \times n$ matrices all of whose entries are either zero or one. How many of these 2^{n^2} matrices have determinant one?

It may be that more time was invested in this problem than in all the others put together. We sat up many long nights, sometimes alone and sometimes in small groups, trying one attack after another. No significant progress was ever made.

One can ask the apparently easier question of how many of these matrices have determinant zero. I once (the night before Yitz Herstein's Algebra exam, in fact) computed the 5×5 case by hand, using a method that filled most of a notebook and didn't seem to generalize. Yitz was anxious to discuss all of this while the exam was in progress, but discussion was finally postponed till later in the afternoon.

A number of related questions were discussed as well. Mike Stob raised the question: What is the largest possible determinant for an $n \times n$ matrix of zeros and ones? At one point, he used the computer at Argonne National Lab to generate 10,000 random 10×10 zero-one matrices and calculate their determinants. To the best of my recollection, the largest determinant that occurred was 24, and it occurred exactly once.

2 The Topology of the Croke Space. The Croke Space C is defined by starting with the Riemann Sphere and identifying points that are roots of the same irreducible polynomial over the rationals. What can be said about the topology of this space? In particular, compute its homology and homotopy groups. (Peter Johnstone observed that C cannot be Hausdorff, since there are irrational square roots of rational numbers arbitrarily close to π, and these are identified with their negatives while π is not.) This problem was posed by Chris Croke. I think it was Kathy Edwards who named the space in question for its inventor.

3. The Locally / Not Problem. This problem came from Peter Johnstone, who was then an instructor at Chicago. I seem to recall that he attributed it to someone else, but I can't remember who. Let P be a property of topological spaces, and say that a space is "locally P" if each point

has a base of neighborhoods all having property P. Now choose a particular property, such as connectedness. Consider successively the class of all spaces that are

(a) connected
(b) locally connected
(c) not locally connected
(d) locally not locally connected
(e) not locally not locally connected

et cetera.

How many of these classes are distinct?

One can also consider the class of all spaces that are

(a′) connected
(b′) not connected
(c′) locally not connected
(d′) not locally not connected

and so forth, and one can ask the same question. Now repeat with connected replaced by compact, or Hausdorff, or noetherian, or... .

Just visualizing some of these classes is already a somewhat ambitious project. As Peter observed at the time, it is not at first obvious what it means for a space to be locally not locally not connected.

4. The Exam Question. Let m, n and k be positive integers. What is the cardinality of the smallest set having m subsets of cardinality n, no two of which have more than k elements in common?

This problem was also posed by Mike Stob, who had a practical application for it: he wanted to give each of 15 students in his Calculus class a personalized take-home exam with 20 problems, and to minimize cheating he wanted no two exams to have more than 5 problems in common. How many problems would he have to make up?

Lauren Feinstone, who was then a graduate student in economics, devised an algorithm which appeared to yield a good upper bound. Unfortunately, computing this bound for the particular values of m, n and k that Mike was interested in turned out to be impracticably complicated.

5. The Word Problem. This was the granddaddy of them all. It was inherited from the more advanced graduate students, in whose lives it had played the same role that zero-one matrices played in ours.

Consider the free group on twenty-six letters $A, B, C, \ldots, Z$. Mod out by the relation that defines two words to be equivalent if (a) one is a permutation of the other and (b) each appears as a legitimate English word in the dictionary. (Obviously this definition may not be invariant under change of dictionary. I hereby arbitrarily and immutably standardize it by naming [**2**] as the unconditional arbiter.) Identify the center of this group.

The fourth floor of Eckhart Hall once housed an enormous chart with 26 rows and columns. In the (i, j)-spot was recorded a proof that letter i commutes with letter j. Some spots were blank, but most were filled in. Several individual letters had been shown to lie in the center. On the other hand, it is very unlikely that the group is actually abelian: it is difficult to imagine a proof that Q commutes with X, for example.

I don't know what became of that chart. If it's been lost, a lot of effort and a lot of history are lost with it.

Other problems that we worked on were less outré and must have been considered by many people besides ourselves. We used to try to count both the number of distinct topologies and the number of distinct homeomorphism types of topologies on a set of n points. We once killed an

afternoon constructing a non-zero polynomial over the quaternions all of whose values are zero (I think we had some help from Kaplansky), and asking exactly which skew fields it's possible to do this for. We also worked on the Riemann Hypothesis and the Poincaré Conjecture, without notable result.

(I fondly remember when Chris Croke proved the Poincaré Conjecture, modulo two facts that seemed plausible. He showed me the proof, we couldn't find anything wrong with it, and we went to ask Dick Lashof about the status of the two "facts". His response: "If those things were true, you could prove the Poincaré Conjecture".)

It's also true that each problem had an associated meta-problem: concoct a plausible scenario in which a mathematician would need to know the answer to this problem. It was generally our conviction that none of the meta-problems was solvable, but I remember a conversation in which Jon Alperin attacked them all with vigor. As I kept challenging him with more and more contrived and outlandish problems, he kept managing to argue that this, too, was a problem whose solution might someday be valuable. I finally stumped him with the Word Problem. He had to admit that it was inconceivable that this one could have any significance for anything. But I think he liked the problem, anyway.

References

1. R. Daniel Mauldin, ed., The Scottish Book: Mathematics Problems from the Scottish Café, Birkhauser, 1981.
2. Webster's Ninth New Collegiate Dictionary, Merriam-Webster, 1985.

MAXIMUM PRODUCTS AND $\lim\left(1+\frac{1}{n}\right)^n = e$

T. N. T. GOODMAN
Department of Mathematical Sciences, The University, Dundee DD1 4HN, Scotland

In a recent edition of this MONTHLY [1] there appears an article by C. W. Barnes entitled "Euler's constant and e" in which he deduces the existence of Euler's constant from the fact that $\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = e$ or, to be more precise, from the inequality

$$e\Big/\left(1+\frac{1}{n}\right) < \left(1+\frac{1}{n}\right)^n < e. \tag{1}$$

He also gives a simple and 'natural' proof of (1), asserting that there is a dearth of such proofs in the textbook literature. We give here a very simple proof of (1) that has a motivation that we think is natural and interesting.

We first give the bare bones of the proof and later fill it out with the flesh of motivation. As in Professor Barnes' proof, the only properties we use of e and the logarithmic function are $\ln e = 1$, $\frac{d}{dx}(\ln x) = \frac{1}{x}$, and $\ln x^n = n \ln x$ for $n = 1, 2, \ldots$.

Let $f_n(t) = t\ln(ne/t)$, $t > 0$. Then $f_n'(t) = \ln(ne/t) - 1$. So f_n has its maximum value at $t = n$. Thus

$$f_n(n+1) < f_n(n), \quad \text{i.e.,} \quad \left(\frac{ne}{n+1}\right)^{n+1} < \left(\frac{ne}{n}\right)^n, \quad \text{i.e.,} \quad e\Big/\left(1+\frac{1}{n}\right) < \left(1+\frac{1}{n}\right)^n.$$

Also

$$f_{n+1}(n) < f_{n+1}(n+1), \quad \text{i.e.,} \quad \left(\frac{(n+1)e}{n}\right)^n < \left(\frac{(n+1)e}{n+1}\right)^{n+1}, \quad \text{i.e.,} \quad \left(1+\frac{1}{n}\right)^n < e.$$

The motivation for this proof is the following problem: given a number $X > 1$, how can we

CENTER SECTION
(Vol. 93, No. 8, October 1986)

Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T:	Textbook	P:	Professional Reading	1-4:	Semesters
C:	Computer Software	L:	Undergraduate Library	** :	Special Emphasis
S:	Supplementary Reading	13:	Grade Level	?? :	Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the Monthly.

General, P, L. Proceedings of the Nineteenth Nordic Congress of Mathematics, Reykjavík 1984. Ed: Jón R. Stefánsson. Icelandic Math Soc, 1985, 270 pp, $20. Contains 9 survey articles and 15 specialized papers, covering all areas of modern mathematics. From Iceland to Finland, mathematics is alive and flourishing. TAV

General, L.** Handbook of Mathematics. I.N. Bronshtein, K.A. Semendyayev. Transl: K.A. Hirsch. Van Nostrand Reinhold, 1985, xv + 973 pp, $37.95 (P). [ISBN: 0-442-21171-6] English translation of a 1978 East German revision of a 1957 Russian handbook. Really a mini-encyclopedia of undergraduate mathematics emphasizing analysis (from calculus to functional analysis), and applied mathematics (statistics, operations research, numerical techniques, information processing). Each chapter sets forth basic definitions, theorems, techniques, and formulas. Nearly 1000 pages of very condensed type. LAS

Elementary, S(13). Mathematics for Pre-college Students. I. Baranov, G. Bogatyrev, O. Bokovnev. MIR (US Distr: Imported Pub), 1985, 376 pp, $9.95. [ISBN: 0-8285-3043-2] Thorough and compact presentation for a "profound reviewing" of arithmetic, algebra and geometry; an inexpensive wealth of information for the serious student; revised from 1983 Russian edition. JNC

Mathematics Appreciation, T(13-14: 1). Mathematical Ideas, Fifth Edition. Charles D. Miller, Vern E. Heeren. Scott Foresman, 1986, 751 pp, $27.95. [ISBN: 0-673-18276-2] Improved edition (Second Edition, TR, March 1974; Third Edition, TR, August-September 1978) with more algebra-based problems throughout, plus an additional chapter on algebra. Each chapter is independent as possible with applications. Useful for teacher training. Well organized and more understandable in areas of logic, counting, and probability. GF

Precalculus, S(13), L. High-School Mathematics. Ed: G.N. Yakovlev. MIR (US Distr: Imported Pub), 1984, $7.95 each. Part One, 327 pp [ISBN: 0-8285-2892-6]; Part Two, 445 pp. [ISBN: 0-8285-2893-4] A synopsis of pre-university mathematics (sets, equations, functions, plane and solid geometry, introduction to calculus) accompanied by over 2000 problems, one-third with complete solutions, compiled by university mathematicians in Moscow from regular written correspondence they maintain with high school students across the Soviet Union. Includes sample mathematics university entrance examinations. LAS

Precalculus, T*(13), S, C, L.** An Introduction to Pascal and Precalculus: Apple Pascal Version. Marvin Marcus. Comput. & Math. Ser. Computer Science Pr, 1986, xix + 421 pp, $29.95 (P). [ISBN: 0-88175-009-3] An alternative to the remedial semester in college algebra and elementary functions, integrating an introduction to Pascal with an exposition of the standard precalculus topics. This spiral bound text/workbook is designed for self study by University of California students. Companion diskette, duplicable for the class; requires Apple UCSD Pascal (graphics and string features exploited). Unique, intriguing. RB

Precalculus, T(13: 1). Algebra for College Students: An Intermediate Approach, Third Edition. Max A. Sobel, Norbert Lerner. Prentice-Hall, 1986, xiv + 511 pp, $28.95. [ISBN: 0-13-021668-2] This new edition (First Edition, TR, August-September 1976) has been rewritten for clarity with significantly increased illustrative examples and applications as well as the number and kind of exercises. Contains additional material in solving linear systems, using matrices, and non-linear systems. More on conic sections and the logarithm. The layout of the book is well designed for use in independent study, with material for self-check, review and testing. GF

Precalculus, T(13: 1). Trigonometry. Howard A. Silver. Prentice-Hall, 1986, xi + 276 pp, $28.95. [ISBN: 0-13-930801-6] Traditional trigonometry text with emphasis on the intuitive right-angle approach. Good examples accompanied by flow chart descriptions of the process used, noting common errors. Easy-to-ready layout and many problems. Built-in use of calculator to simplify and rein-

force concepts. GF

Precalculus, T(13: 1, 2). Fundamentals of College Algebra, Second Edition. Charles D. Miller, Margaret L. Lial. Scott Foresman, 1986, 487 pp, $26.95 [ISBN: 0-673-18242-8]; Algebra and Trigonometry, Fourth Edition, 1986, 706 pp, $28.95 [ISBN: 0-673-18296-7]; Algebra and Trigonometry, Fourth Edition, An Alternate Approach, 1986, 706 pp, $28.95. [ISBN: 0-673-18298-3] College Algebra covers traditional topics with abundant exercise sets. Important definitions "boxed" for emphasis. Problem solving addressed through general suggestions for word problems as well as highlighting procedures for specific applications. Algebra and Trigonometry contains same material and features, plus three trigonometry chapters (unit circle approach); the Alternate Approach uses the right triangle approach. (For TRs of earlier editions, see October 1978, December 1980, October 1982, and October 1983.) RD

Finite Mathematics, T(13: 1). Mathematics for Business and Economics. Robert H. Nicholson. McGraw-Hill, 1986, xx + 524 pp, $32.95. [ISBN: 0-07-046491-X] Techniques from systems of linear equations and inequalities, matrix algebra, linear programming, mathematics of finance, calculus, and probability are presented in a dreary format with almost no illustrations to break the monotony; definitions, theorems, etc., aren´t clearly distinguished. JNC

Education, P. Logo in the Schools. Ed: Cleborne D. Maddux. Haworth Pr, 1985, 305 pp, $14.95 (P). [ISBN: 0-86656-425-X] An excellent compilation of articles spanning three critical issues: 1) existing LOGO research; 2) examples and reports of LOGO projects; and 3) classroom uses for LOGO. Many program samples and references listed throughout. RD

Education, P. Evaluating Intervention Programs: Applications from Women´s Programs in Math and Science. Barbara Gross Davis, Sheila Humphreys. Teacher College Pr, 1985, xvi + 228 pp, $15.95 (P). [ISBN: 0-8077-2787-3] A handbook for developing effective evaluation programs for educational projects, especially for intervention programs designed to interest young women in scientific courses. Revision of a 1983 version titled Evaluation Counts. LAS

History, S, P, L. Studies in the Exact Sciences in Medieval Islam. Ali A. Al-Daffa, John J. Stroyls. Wiley, 1984, x + 243 pp, $39.95. [ISBN: 0-471-90320-5] Seven essays by the authors on the development and transmission of Islamic and Arabic science and technology during the Middle Ages. East-West transmission during the Crusades; myths about the Pythagorean theorem, logarithms; 15th century work on Euclid´s parallel postulate; mathematical work of Ibn Sina; early use of numerical methods; medieval Middle Eastern theories of equations. RB

Logic, T(16-18: 1), P, L. Building Models by Games. W. Hodges. London Math. Soc. Stud. Texts, V. 2. Cambridge U Pr, 1985, vi + 311 pp, $42.50; $16.95 (P). [ISBN: 0-521-26897-4; 0-521-31716-9] Text for a model theory course which assumes familiarity with cardinals, ordinals, logical notation and abstract algebra. Uses two-person games to describe different kinds of forcing. Includes applications to algebra. Exercises range from drill to research. List of open questions and extensive bibliography. KS

Logic, P?. Logic as Grammar. Norbert Hornstein. MIT Pr, 1984, xi + 176 pp, $20. [ISBN: 0-262-08137-7] A book on philosophy of linguistics devoted to meaning in natural language. Current model-theoretic semantical approaches to meaning are rejected as explanations of a native speaker´s interpretive abilities, in favor of a generative grammar approach. Of interest to linguists, perhaps thereby to artificial intelligence specialists in natural language; no direct mention of computing or mathematical logic. RB

Logic, P. Recent Developments in Quantum Logic. Ed: Peter Mittelstaedt, Ernst-Walther Stachow. Grund. der exakten Naturwissenschaften, B. 6. Bibliographisches Institut, 1985, 352 pp. [ISBN: 3-411-01695-7] The proceedings (twenty-three papers) of the International Symposium on Quantum Logic held at the University of Cologne, Germany, June 13-16, 1984. Contributors include physicists, mathematicians, and philosophers. The symposium was concerned with the formal language of quantum logic, its logic and semantics, the relation between these linguistic structures and mathematical models, and the interpretation of quantum physics. LCL

Foundations, P. Beyond Analytic Philosophy: Doing Justice to What We Know. Hao Wang. MIT Pr, 1986, xii + 273 pp, $17.50. [ISBN: 0-262-23124-7] A critical review of the Russell-Carnap-Quine tradition of Anglo-American analytic philosophy, with suggestions for a new approach inspired in part by the "objectivism" approach of Kurt Gödel. LAS

Foundations, T(13-15: 1), S, L. Number Systems and the Foundations of Analysis. Elliott Mendelson. Robert E Krieger, 1985, xii + 358 pp, $24.95. [ISBN: 0-89874-818-6] A reprint of the original 1973 edition, with corrections. The aim is to study the basic number systems of mathematics, beginning with an axiomatic presentation of the natural numbers, and proceeding by set-theoretical methods to the construction of the integers, the rationals, and the reals by Cauchy sequences. Includes an appendix on Dedekind cuts. LCL

Foundations, T(18), S, P. Boolean-Valued Models and Independence Proofs in Set Theory. J.L. Bell. Logic Guides, V. 12. Oxford U Pr, 1985, xx + 165 pp, $22.95. [ISBN: 0-19-853241-5] This edition, reset in TEX, incorporates an account of iterated Boolean extensions and the consistency and independence of Souslin´s hypothesis, as well as several other substantial modifications in the original text (First Edition, TR, June-July 1978). LCL

Combinatorics, T(17: 2), S, P, L. Theory of Codes. Jean Berstel, Dominique Perrin. Pure & Appl. Math., V. 117. Academic Pr, 1985, xiv + 433 pp, $60. [ISBN: 0-12-093420-5] A systematic and comprehensive study of the theory of variable length codes. Chapter titles include prefix codes, biprefix codes, automata, groups of biprefix codes, densities, conjugacy and the polynomial of a code. Encoding, as treated here, can be formulated as the study of embeddings of one free monoid into another. Includes exercises and references. CEC

Combinatorics, S(17-18), P. Lecture Notes in Mathematics-1168: Hadamard Matrices and Their Applications. S.S. Agaian. Springer-Verlag, 1985, 227 pp, $14.40 (P). [ISBN: 0-387-16056-6] While title refers to applications, two-thirds of this volume is devoted to construction of Hadamard matrices, many methods of which are of recent vintage or are new. Last quarter of the book treats applications to information theory, designs and other combinatorial matters. SS

Discrete Mathematics, T(17-18), S, P. Design Theory. Thomas Beth, Dieter Jungnickel, Hanfried Lenz. Bibliographisches Inst, 1985, 688 pp. [ISBN: 3-411-01675-2] Probably the most comprehensive textbook on the subject. Assuming no background in design theory, authors cover all the main concepts and explore interesting applications and interrelationships with algebra, geometry, and other areas. SS

Discrete Mathematics, T*(13-14: 1), S. An Introduction to Discrete Mathematics. Steven Roman. Saunders College, 1986, xiii + 455 pp, $32.95. [ISBN: 0-03-064019-9] Topics (sets, logic, relations, combinatorics, and graphs) are presented with a freshness and vitality that is within the spirit of modern discrete mathematics. The content, examples, and exercises, are motivated by the presence of the computer, but this is a mathematics book: the student is expected to handle definition, theorem, and proof. Comfortably paced; ample explanation and motivation; attractive format; instructive examples; large exercise sets (many on the very-easy side). LCL

Number Theory, P. 1000 Problems. A.S. Moiseenko (10-12 Kimball St., Belleville, NJ), 1985, 413 pp, (P). Sample solutions to $18^2 + 26^2$ Diophantine equations. Example: $x_1^3 + x_2^3 + \ldots + x_7^3 = y^3$ has a solution (5,12,24,35,48,73,96;113). There are no words in this book![1] BC

Number Theory, T(17-18: 1, 2), S, P. The Arithmetic of Elliptic Curves. Joseph H. Silverman. Grad. Texts in Math., V. 106. Springer-Verlag, 1986, xii + 400 pp, $48. [ISBN: 0-387-96203-4] Elliptic curves bridge the enormous gap between Pythagorean triples and Faltings' theorem, and possess many beautiful, even amazing properties of their own. This excellent introduction, emphasizing an algebraic approach, will attract many new mathematicians to the subject. A good companion to Koblitz' analytic treatment (GTM 97, TR, April 1985) of elliptic curves and modular forms. BC

Number Theory, T(14: 1), S, L*. Lectures on Number Theory. Adolf Hurwitz, Nikolaos Kritikos. Transl: William C. Schulz. Universitext. Springer-Verlag, 1986, xiv + 273 pp, $19.80 (P). [ISBN: 0-387-96236-0] A solid introduction to elementary number theory. Considering that the text is based on lecture notes that are more than 70 years old, this is an amazingly clear and readible presentation. The translator has added a good collection of exercises. CEC

Number Theory, P. Class Field Theory. Jürgen Neukirch. Grund. der math. Wissenschaften, B. 280. Springer-Verlag, 1986, viii + 140 pp, $29.50. [ISBN: 0-387-15251-2] Class field theory without cohomology, based instead on a group-theoretic "class field axiom." Self-contained, assuming Galois and algebraic number theory. Terse and to the point. BC

Linear Algebra, T(14-15: 1), L. Introduction to Linear Algebra with Applications. Stephen H. Friedberg, Arnold J. Insel. Prentice-Hall, 1986, xiv + 463 pp, $30.95. [ISBN: 0-13-485988-X] A good blend of theory, technique, and application, with exercises of each type. A social-science slant to many applications. Appendix includes seven BASIC programs. Diskette available. BC

Linear Algebra, T*(14: 1). Elementary Linear Algebra. Richard O. Hill, Jr. Academic Pr, 1986, ix + 404 pp, $25. [ISBN: 0-12-348460-X] Emphasizes applications and numerical considerations required in computer implementation; topics motivated via examples, e.g., the least squares problem introduces orthogonality; gives option of including a minimal presentation of determinants before eigenvalues; Instructor's Manual available. JNC

Group Theory, P. Group Theory: Essays for Philip Hall. Ed: K.W. Gruenberg, J.E. Roseblade. Academic Pr, 1984, vii + 353 pp, $65. [ISBN: 0-12-304880-X] Commissioned essays surveying areas in which Hall made important contributions, originally intended to mark his 80th birthday. Finite non-solvable groups, finite and finitely presented soluble groups, nilpotent groups, classification of torsion free nilpotent groups, group rings of polycyclic groups, the algebra of partitions; authors include Thompson and Robinson. RB

Group Theory, T(18), S, P. Group Theory II. Michio Suzuki. Grund. der math. Wissenschaften, B. 248. Springer-Verlag, 1986, x + 621 pp, $89. [ISBN: 0-387-10916-1] A translation of chapters four through six of the author's book Gunron, covering the theory of commutators and the theory of finite groups. LCL

Group Theory, P. Contributions to Group Theory. Ed: Kenneth I. Appel, John G. Ratcliffe, Paul E. Schupp. Contemp. Math., V. 33. AMS, 1984, xi + 519 pp, (P). [ISBN: 0-8218-5035-0] Expository articles and invited research papers honoring Roger Lyndon's 65th birthday. The editors, MacLane and Keisler contribute a biography, write on Lyndon's dissertation and his contributions to cohomology of groups, mathematical logic, and group theory; 27 research papers cover various fields, especially combinatorial group theory. RB

Group Theory, S(18), P. Computational Group Theory. Ed: Michael D. Atkinson. Academic Pr, 1984, xii + 373 pp, $60. [ISBN: 0-12-066270-1] The proceedings of a symposium which took place from July 30 to August 9, 1982 at Durham University. Includes contributions from five key speakers: Cannon, Conway, Leach, Neubüser, and Newman along with 23 other papers. CEC

Topological Groups, P. Harish-Chandra Homomorphisms for p-adic Groups. Roger Howe, Allen Moy. CBMS Reg. Conf. Ser. in Math., No. 59. AMS, 1985, xi + 76 pp, $13 (P). [ISBN: 0-8218-0709-9] Notes based on lectures given in Chicago, 1983. The Harish-Chandra homomorphism is important for representation theory of semi-simple groups. Howe and Moy are developing a p-adic analogue, here for GL_n, using Hecke algebras. BC

Algebra, P. Universal Algebra and Its Links with Logic, Algebra, Combinatorics and Computer Science. P. Burmeister, et al. Res. & Expos. in Math., V. 4. Heldermann Verlag, 1984, vii + 243 pp, $62 (P). [ISBN: 3-88538-204-0] Proceedings of a 1983 Darmstadt workshop on universal algebra. JD-B

Calculus, T(13-14: 2, 3). Technical Mathematics with Calculus. C.E. Goodson, S.L. Miertschin. Wiley, 1985, xv + 1146 pp, $33.95. [ISBN: 0-471-86639-3] College algebra, trigonometry and a term of calculus for college-level technical, pre-engineering and career-oriented programs. Informal presentation, special attention to word problems including sections devoted to specific applications, appendices on calculator and computer, chapter on probability and statistics. RB

Calculus, T(13: 2), S, L. The Calculus Tutoring Book. Carol and Robert B. Ash. IEEE Pr, 1985, x + 534 pp. [ISBN: 0-87942-183-5] Short, informal, catchy, intuitive introduction to elementary calculus. No pretense of rigor is made; authors cites "formalism" in standard texts as obstacle to learning. Much is omitted (epsilons, deltas, mean value theorem, almost all proofs) but standard topics appear. Geometric intuition is emphasized; curiously, numerical viewpoint is not. Baby gone with bathwater? Worth a look to decide. Note publisher. PZ

Calculus, T*(13: 1). Calculus for Management, Social and Life Sciences. Dennis D. Berkey. Saunders College, 1986, xii + 419 pp, $31.95. [ISBN: 0-03-004764-1] A nice one-semester treatment of calculus emphasizing applications, primarily in business and economics, with numerous graphical illustrations; contains an index of applications, word problems in nearly every section, and optional references to computers in both text and exercises; Student's Solution Manual and Instructor's Manual available. JNC

Calculus. A First Course in Calculus, Fifth Edition. Serge Lang. Undergrad. Texts in Math. Springer-Verlag, 1986, xv + 726 pp, $39.95. [ISBN: 0-387-96201-8] A quick (1/2 usual length), informal (but not unsophisticated) treatment of elementary calculus, omitting multiple integrals, differential equations, conic sections, numerical integration. Most proofs, all epsilon-delta material are in appendices--limits and continuity are treated intuitively. Taylor's theorem with remainder calculations precedes and motivates series discussion. Lively, active writing style. An unusual, interesting alternative to standard texts. (Third Edition, TR, February 1974; Fourth Edition, June-July 1978.) PZ

Calculus, T(13: 2, 3). Calculus with Analytic Geometry, Third Edition. Robert Ellis, Denny Gulick. Harcourt Brace Jovanovich, 1986, xi + 1060 pp, $40.95. [ISBN: 0-15-505735-5] New to this edition (First Edition, TR, August-September 1978) are cumulative review exercises in chapters 3-15, an algorithmic approach to bisection and Newton-Raphson methods; computer-drawn figures in chapter on partial derivatives. JNC

Calculus, T(13: 1, 2). Technical Mathematics and Calculus. Rudolph E. Lynn. Wiley, 1985, xv + 963 pp, $33.95. [ISBN: 0-471-87902-9] A text intended primarily for students in vocationally-oriented engineering technology programs, covering college algebra, trigonometry and first-term calculus. Some features: conversational tone, fully-detailed examples, emphasis on checking work, engineering graphs, calculator programming. RB

Complex Analysis, T(18: 1), P. Complex Manifolds and Deformation of Complex Structures. Kunihiko Kodaira. Transl: Kazuo Akao. Grund. der math. Wissenschaften, B. 283. Springer-Verlag, 1985, x + 465 pp, $48. [ISBN: 0-387-96188-7] An introduction to the theory developed since the fifties (though with roots dating back to Riemann) by the author and D.C. Spencer. The first half is a general introduction to analysis and geometry on complex manifolds. A long appendix treats elliptic partial differential operators on manifolds. PZ

Differential Equations, T(17-18), P. Stability and Periodic Solutions of Ordinary and Functional Differential Equations. T.A. Burton. Math. in Sci. & Eng., V. 178. Academic Pr, 1985, x + 337 pp, $65. [ISBN: 0-12-147360-0] A textbook (second-semester graduate course) and reference. The aim is for unity and cohesiveness, especially with regard to the applicability of techniques of ordinary differential equations to functional differential equations. LCL

Numerical Analysis, T(15-16: 1), S, L. Mathematics and CAD, Volume 1: Numerical Methods for CAD. Yvon Gardan. MIT Pr, 1986, 165 pp, $25. [ISBN: 0-262-07097-9] Basic mathematical concepts useful to the computer-aided-design systems designer. Many references to the literature. Topics: basic problems using graphics (geometric transformations, projections and perspectives), curves and surfaces, numerical methods for solving linear and non-linear equations, the finite element method. DFA

Functional Analysis, P. Lecture Notes in Mathematics-1162: Trajectory Spaces, Generalized Functions and Unbounded Operators. S.J. L. van Eijndhoven, J. de Graaf. Springer-Verlag, 1985, iv + 272 pp,

$17.60 (P). [ISBN: 0-387-16065-5] This book presents a new model for generalized function theories. Using a Hilbert space and a self-adjoint operator as initial data, the book outlines a method for constructing a space of test functions, a space of generalized functions, and dual pairing. This means that the topology of the space of generalized functions is defined by the Hilbert space instead of by a topology placed on the space of test functions. The authors state that this permits a systematic exposition of topological tensor products and kernel theorems. AM

Functional Analysis, T(17: 1, 2), S, P. Convexity Methods in Variational Calculus. Peter Smith. Appl. & Eng. Math. Ser., V. 1. Wiley, 1985, x + 222 pp, $41.95. [ISBN: 0-471-90679-4] Integrals that arise in the calculus of variations are often convex, and therefore admit useful extremum principles. This monograph develops convexity methods in the general setting of functionals on normed linear spaces and applies them mainly to functionals arising from operator equations in applied mathematics. Intuitive style; first chapter introduces basic notions at elementary level. With exercises, many solved examples. PZ

Analysis, P. An Introduction to Minimal Currents and Parametric Variational Problems. Enrico Bombieri. Math. Reports, V. 2, Part 3. Harwood Academic, 1985, vii + 99 pp, $25 (P). [ISBN: 3-7186-0299-7] Divided into five sections: currents and varifolds (fundamental notions, results, examples), isoperimetric inequalities and variational formulas, variational methods applicable to some problems on currents, the minimal surface problem, and the Bernstein problem. LCL

Analysis, P. Eleven Papers in Analysis. P.L. Shabalin, et al. AMS Transl., Ser. 2, V. 127. AMS, 1986, v + 119 pp, $48. [ISBN: 0-8218-3094-5]

Analysis, P. Function Spaces on Subsets of R^n. Alf Jonsson, Hans Wallin. Math. Reports, V. 2, Part 1. Harwood Academic, 1984, xiv + 221 pp, $48 (P). A systematic treatment including Sobolev, Lipschitz, and Besov spaces, imbedding theorems and trace space analysis. Extensive bibliography. TAV

Analysis, P. Multiple Gaussian Hypergeometric Series. H.M. Srivastava, Per W. Karlsson. Ser. in Math. & Its Applic. Halsted Pr, 1985, 425 pp, $74.95. [ISBN: 0-470-20100-2] A complete classification and convergence theory exists for Gaussian hypergeometric series in one and two variables. This monograph solves the same problems in dimension three, and systematically treats higher-dimensional cases. A long chapter suggests research problems. Extensive bibliography. PZ

Differential Geometry, P. Equilibrium Capillary Surfaces. Robert Finn. Grund. der math. Wissenschaften, B. 284. Springer-Verlag, 1986, xvi + 245 pp, $57. [ISBN: 0-387-96174-7] Liquid surfaces offer a wonderful opportunity for existence theorems and differential geometry. New technology has rekindled interest in capillarity, studied extensively in the nineteenth century (Laplace and Gauss being notable contributors, as usual). Chapters here on capillary tubes, sitting and hanging drops, and what happens without gravity. BC

Differential Geometry, S(16-18), P. Dynamical Systems: A Differential Geometric Approach to Symmetry and Reduction. Giuseppe Marmo, et al. Wiley, 1985, xii + 377 pp, $64.95. [ISBN: 0-471-90339-6] Exposition of the foundations of mechanics within the framework of differential geometry. Includes introductory discussions of manifolds, tangent and cotangent bundles, foliations, Lie groups and Lie algebras and their actions on manifolds. Also describes the Hamiltonian and Lagrangian formalisms for mechanics, Noether's theorem, symplectic geometry, and the use of Lie groups to analyze the underlying symmetries of problems in mechanics. AM

Differential Geometry, P. Eigenvalues in Riemannian Geometry. Isaac Chavel. Pure & Appl. Math., No. 115. Academic Pr, 1984, xiv + 362 pp, $62. [ISBN: 0-12-170640-0] Introduction to the geometry of the Laplace operator: basic techniques and results in a field in which topology and Riemannian geometry interact with partial differential equations, probability, number theory. The Laplacian acting on functions; eigenvalues of the Laplacian for compact manifolds and domains with compact closure; associated questions concerning the heat equation. RB

Differential Geometry, P. Lecture Notes in Mathematics-1157: Classifying Immersions into R^4 over Stable Maps of 3-Manifolds into R^2. Harold Levine. Springer-Verlag, 1985, v + 163 pp, $12 (P). [ISBN: 0-387-15995-9] For maps f and h from a compact, oriented 3-manifold M into R^2, a regular homotopy of immersions (f,h) of M into $R^2 \times R^2$ is a homotopy such that each level (f,h_t) is an immersion. These lecture notes classify immersions (f,h) up to regular homotopy in the case where f is stable. RB

Geometry, S(16-18), P. Finite Geometries. Ed: Catharine Anne Baker, Lynn Margaret Batten. Lect. Notes in Pure & Appl. Math., V. 103. Dekker, 1985, xi + 375 pp, $69.75 (P). [ISBN: 0-8247-7488-4] From the July 1984 conference on finite geometry, held in Winnipeg, a collection of 31 articles--some survey papers but most research papers--ranging over a variety of questions concerning finite geometries and related combinatorial matters. SS

Geometry, S*, P, L.** Results and Problems in Combinatorial Geometry. V.G. Boltjansky, I. Ts. Gohberg. Cambridge U Pr, 1985, 108 pp, $29.95; $9.95 (P). [ISBN: 0-521-26298-4; 0-521-26923-7] An inviting discussion of a few problems from the theory of convex bodies, mainly connected with the partition of a set into smaller parts. The presentation is accessible to beginners; most of the discussion is restricted to two and three dimensions, but the book brings the reader up-to-date with references and notes to the state-of-the-art. Includes a collection of unsolved and partially-solved problems that even novices can understand and speculate about. LCL

Geometry, S*(13-15), L. Images of Geometric Solids. N.M. Beskin. Transl: Valery Barvashov. Little Math. Lib. MIR (US Distr: Imported Pub), 1985, 78 pp, $2.95 (P). [ISBN: 0-8285-3028-9] Another in the series of Soviet supplementary pamphlets intended to introduce upper secondary students to interesting advanced mathematics. This one develops the theory of drawing three-dimensional solids in the plane as a compromise betweeen obvious resemblance and faithful measurement. An excellent excursion into an area of applied geometry closely related to uses in computer graphics. LAS

Topology, P. Lecture Notes in Mathematics-1167: Geometry and Topology. Ed: J. Alexander, J. Harer. Springer-Verlag, 1985, vi + 292 pp, $17.60 (P). [ISBN: 0-387-16053-1] Papers submitted by participants in the 1983-84 Special Year in Topology at the University of Maryland covering five themes: 3-manifolds and hyperbolic geometry; homological methods; 4-manifolds; geometric and analytic methods; and growth of groups. LAS

Optimization, T(15-17), L. Linear and Combinatorial Programming. Katta G. Murty. Robert E Krieger, 1985, xxiii + 567 pp, $41.50. [ISBN: 0-89874-852-6] Unaltered reprint of the 1976 original edition (TR, August-September 1977), without any updating even of references (section on "available computer programs" contains only citations from 1973-75). Despite its age, it contains a wealth of appealing and practical applications of linear and integer programming problems, together with network algorithms, travelling salesman, and complementarity problems. LAS

Probability, P, L*. Problems in the Theory of Probability. B.A. Sevastyanov, V.P. Chistyakov, A.M. Zubkov. Transl: Irene Aleksanova. MIR (US Distr: Imported Pub), 1985, 236 pp, $7.95. [ISBN: 0-8285-3044-0] A useful collection of problems--mostly theoretical--on topics including geometric probability, limit theorems, Markov chains. At this price, a must for teachers needing exam questions. All problems have complete solutions. TAV

Probability, P. Parameter Estimation for Stochastic Processes. Yu. A. Kutoyants. Res. & Expos. in Math., V. 6. Transl: B.L.S. Prakasa Rao. Heldermann Verlag, 1984, viii + 206 pp, $56 (P). [ISBN: 3-88538-206-7] A translation of the 1980 work in Russian. Maximum likelihood and Bayes estimates are developed for parameters associated with Gaussian, diffusion and non-homogeneous Poisson processes. Excellent bibliography, but note the price! TAV

Computer Literacy, T, S(13). Using Computers Today. David R. Sullivan, T.G. Lewis, Curtis R. Cook. Houghton Mifflin, 1986, xxiii + 633 pp, $27.95 (P). [ISBN: 0-395-40639-0] Computer literacy for technical and non-technical students, using personal computing as a vehicle to explain general computing concepts. It is available in three versions (generic, IBM, Apple) that differ only in how they cover BASIC. Good preview for business, engineering, science, and liberal arts students. GF

Computer Literacy, S(13-15), L. The Biology of Computer Life: Survival, Emotion and Free Will. Geoff Simons. Birkhauser Boston, 1985, xii + 236 pp, $11.95 (P). [ISBN: 0-8176-3299-9] An imaginative sequel to the author's 1983 Are Computer's Alive? (TR, February 1985) hawking an affirmative answer--a world in which homo sapiens and machine sapiens converge in their attributes of reproduction, autonomy, free will, emotion, even sexuality. Speculative, interesting, occasionally prolix; well-referenced to appropriate AI literature. LAS

Computer Literacy, T(13-14: 1, 2). Computers and Information Systems. Jerome S. Burstein. Holt, Rinehart & Winston, 1986, xxi + 694 pp, $29.95. [ISBN: 0-03-070519-3] Intended for introductory course on information systems. Various units explore hardware, software, system, and social issues. Highly readable with chapter outline, objectives, summary, and review aids. Enhancements include color photos, graphs, and drawings. Current and comprehensive presentation. RD

Computer Programming, T(13-14: 1). Basic Programming with Structure and Style. Thomas S. Logsdon. Boyd & Fraser, 1985, xvi + 418 pp, (P). [ISBN: 0-87835-808-0] A text for a one-semester course in BASIC programming. Example programs written in Microsoft BASIC on microcomputers; appendices on IBM/PC, Apple IIe, Radio Shack TRS-80/4, timesharing systems; chapters on added Microsoft control structures (e.g., WHILE) and graphics. Stepwise refinement, top-down algorithm development are excluded. RB

Computer Programming, T(13-14: 1). VAX BASIC. Clifford Townsend. Holt, Rinehart & Winston, 1986, 251 pp, $18.95 (P). [ISBN: 0-03-921848-1] Introductory BASIC programming text. Each chapter provides overview, BASIC commands, VAX facts, summary and exercises. A "how-to" approach with problem-solving emphasis. RD

Computer Programming, T(13-14), S. Structured Fortran with WATFIV: Text and Reference, Third Edition. John B. Moore. Reston, 1985, xxiii + 502 pp, (P). [ISBN: 0-8359-7118-X] A textbook/manual for a course in Fortran programming using the WATFIV compiler developed at the University of Waterloo, Ontario. This edition adds terminal-oriented input/output, comparison with Fortran-77, and greater reinforcement of top-down design and stepwise refinement. RB

Computer Programming, T(13: 1). FORTRAN for Technologists and Engineers. James Valentino. Holt, Rinehart & Winston, 1985, xiii + 594 pp, $24.95. [ISBN: 0-03-060569-5] Presupposes only elementary algebra and trigonometry. Complete presentation of ANSI standard FORTRAN and of many features of WATFIV-S and FORTRAN 77. Emphasizes good programming style. Flowcharts. Examples and exercises focus on practical scientific and engineering problems. DFA

Software Systems, P. Nested Transactions: An Approach to Reliable Distributed Computing. J. Eliot B. Moss. MIT Pr, 1985, xi + 160 pp, $20 (P). [ISBN: 0-262-13200-1] Software approach for architec-

ture of distributed information systems to produce reliability in presence of failures. Design based on hierarchies of atomic actions ("transactions"), locking for synchronization. Algorithms for management, synchronization of transactions, recovery, progress, distributed deadlock detection. RM

Software Systems, P. Lecture Notes in Computer Science-203: EUROCAL ´85, V. 1. Ed: Bruno Buchberger. Springer-Verlag, 1985, v + 233 pp, $14.60 (P). [ISBN: 0-387-15983-5] 14 invited papers from the European Conference on Computer Algebra at Linz, Austria, in April 1985. Several papers are introductory, one is historical, and others treat links between computer algebra and neighboring fields. PZ

Software Systems, P. Lecture Notes in Computer Science-204: EUROCAL ´85, V. 2. Ed: Bob F. Caviness. Springer-Verlag, 1985, xvi + 650 pp, $37.50 (P). [ISBN: 0-387-15984-3] A large collection of full papers, extended abstracts, and informal research contributions from the European Conference on Computer Algebra at Linz, Austria, in April 1985. Includes extensive index and listings of papers presented at other recent conferences on computer algebra. PZ

Software Systems, P. Program Evolution: Processes of Software Change. Ed: M.M. Lehman, L.A. Belady. APIC Stud. in Data Proc., V. 27. Academic Pr, 1985, xiii + 538 pp, $40. [ISBN: 0-12-442440-6] Two dozen papers, mostly reprints from other sources, on the evolutionary characteristics of complex programming processes "reminiscent of the evolution of biological organisms and social groupings." The thesis of these authors is that intrinsic pressure for evolution rather than lack of programmer´s foresight is responsible for the software crisis. LAS

Computer Science, P. Computer Architectures for Spatially Distributed Data. Ed: Herbert Freeman, Goffredo G. Pieroni. NATO ASI Ser. F, V. 18. Springer-Verlag, 1985, viii + 391 pp, $65. [ISBN: 0-387-12886-7] Conference proceedings on special parallel architectures (SIMD, pyramids, and extensions) for spatially distributed data (e.g., arising from multidimensional processes and problems). Applications in the areas of image processing, pattern recognition, computer cartography. RM

Computer Science, P. The Second RIKEN International Symposium on Symbolic and Algebraic Computation by Computers. Ed: N. Inada, T. Soma. World Scientific, 1984, viii + 246 pp, $30. [ISBN: 9971-50-021-3] Proceedings of an August 1984 symposium held in Japan. Eighteen papers (some in abstract form) on theory and applications of computer algebra systems. PZ

Computer Science, P. Lecture Notes in Computer Science-210: STACS 86. Ed: B. Monien, G. Vidal-Naquet. Springer-Verlag, 1986, ix + 368 pp, $20.50 (P). [ISBN: 0-387-16078-7] Proceedings of the third symposium on theoretical aspects of computer science in Paris. Variety of contributions from complexity theory, combinatorial algorithms, semantics and correctness, parallelism. RM

Applications, T*(16-17: 2), S*, P*. Space Kinematics and Lie Groups. Adolf Karger, Josef Novák. Gordon & Breach, 1985, xv + 422 pp, $88. [ISBN: 2-88124-023-2] A translation of the 1978 Czech publication. Roughly in two somewhat independent parts, one for those who deal with concrete problems and the other for the theoretically oriented, with an introductory chapter introducing differentiable manifolds and Lie groups. Numerous non-trivial examples to make the point that "even relatively simple motions lead to rather complicated expressions." All formulas are in forms applicable for computer processing. Scattered exercises without solutions. One hundred sixty references, mostly to books and papers in French and German. JK

Applications, P. Lecture Notes in Computer Science-209: Advances in Cryptology. Ed: T. Beth, N. Cot, I. Ingemarsson. Springer-Verlag, 1985, vii + 491 pp, $22.80 (P). [ISBN: 0-387-16076-0] Proceedings of EuroCrypt 84, an annual workshop on cryptology held in April 1984 in Paris. Includes papers on classical methods, public-key systems, number theory, networks, and smart cards--credit cards with built-in microprocessors. LAS

Applications (Artificial Intelligence), P. Machine Learning: An Artificial Intelligence Approach. Ed: Ryszard S. Michalski, Jaime G. Carbonell, Tom M. Mitchell. Morgan Kaufmann. Volume I, 1983, xi + 572 pp, $39.95 [ISBN: 0-935382-05-4]; Volume II, 1986, x + 738 pp, $39.95. [ISBN: 0-934613-00-1] Two tutorial volumes with contributed chapters by leaders in AI covering such themes as learning from examples, from observation and discovery, from instruction, and by analogy. Includes several reports on applications both to science and to pedagogy. A superb survey of contemporary research in machine learning suited to readers without special expertise in AI. LAS

Applications (Astronomy), S*(14-16), P, L.** The Invisible Universe: Probing the Frontiers of Astrophysics. George B. Field, Eric J. Chaisson. Birkhauser Boston, 1985, xiv + 195 pp, $19.95. [ISBN: 0-8176-3235-2] A picture of the new universe of dark objects in space revealed by modern ground and space-born observatories designed to detect radiation invisible to the human eye: such things as interstellar matter and quasars help reveal the large-scale structure of the universe and the fundamental forces of nature. Based on a 1982 report of the National Research Council--astronomy´s "David" report. Illustrated with colorplates and, concluding each chapter, brief personal accounts by the authors. LAS

Applications (Biology), S(15-18), P. The Book of L. G. Rozenberg, A. Salomaa. Springer-Verlag, 1986, xv + 471 pp, $39.50. [ISBN: 0-387-16022-1] The study of "L-systems" dates from 1968 when Aristid Lindenmayer first introduced language-theoretic models for studying developmental biology. This volume, with 39 papers, represents the full range of research in this area; contributors include mathematicians, theoretical computer scientists, and biologists. Two titles give the flavor:

"Hierarchical Aspects of Plant Development," and "On Cyclically Overlap-Free Words in Binary Alphabets." LCL

Applications (Biology), P. Lecture Notes in Biomathematics-63: Temporal-Pattern Learning in Neural Models. Carme Torras. Springer-Verlag, 1985, vii + 227 pp, $16.40 (P). [ISBN: 0-387-16046-9] A monograph studying neurophysiological mechanisms underlying the ability of animals to learn rhythms. Part I, description of experimental and theoretical framework. Part II, detailed model. Part III, analytic study of the model, and Part IV, description of neural network model proposed to study intraneural parameters, initial state and stimulation conditions. GF

Applications (Biology), P. Complexity, Language, and Life: Mathematical Approaches. Ed: John L. Casti, Anders Karlqvist. Biomathematics, V. 16. Springer-Verlag, 1986, xiii + 281 pp, $49.50. [ISBN: 0-387-16180-5] 10 papers prepared as a sequel to a May 1984 workshop on various mathematical views of the evolution of complex systems. Interdisciplinary in nature and innovative by necessity, these papers offer an extraordinary perspective on the mathematization of theories of life. LAS

Applications (Cognitive Science), P. Getting Computers to Talk Like You and Me. Rachel Reichman. MIT Pr, 1985, xiii + 221 pp, $20. [ISBN: 0-262-18118-5] Begins with an analysis of person-to-person verbal communication. The results of this analysis are then incorporated into a computer program that keeps track of the flow of a conversation and checks for its coherence. The final but as-yet-remote goal is the development of a computer program that can understand everyday discourse. SG

Applications (Management), P. Analytical Planning: The Organization of Systems. Thomas L. Saaty, Kevin P. Kearns. Intern. Stud. in Modern Appl. Math. & Comp. Sci., V. 7. Pergamon Pr, 1985, viii + 208 pp, $40. [ISBN: 0-08-032599-8] A mostly non-mathematical discourse on planning, including, however, the eigenvector-based analytical hierarchy process by which different options can be compared in a multi-objective setting. LAS

Applications (Physics), S(16-18), P. Lecture Notes in Engineering-15: The Shallow Water Wave Equations: Formulation, Analysis and Application. I. Kinnmark. Springer-Verlag, 1985, xxv + 187 pp, $18 (P). [ISBN: 0-387-16031-0] Describes flow in well-mixed fluids experiencing tidal and atmospheric forcing. After extensive analysis, the results are compared to empirical data of several well-documented examples, e.g., the North Sea. The careful mathematical analysis is imbedded in ample lucid discussion. MU

Applications (Physics), T(13: 1), S, L. Elements of Relativity Theory. D.F. Lawden. Wiley, 1985, x + 108 pp, $14.95 (P). [ISBN: 0-471-90852-5] Presuming nothing more than good algebraic skills, this carefully written text covers much of special relativity and introduces the general theory. The discussion is thorough and lucid, sufficient to guide a dedicated novice through the carefully chosen equations, and prepare him or her to solve the "practical" problems posed at the ends of the chapters. A good buy for any undergraduate library. MU

Applications (Physics), P. On Growth and Form: Fractal and Non-Fractal Patterns in Physics. Ed: H. Eugene Stanley, Nicole Ostrowsky. NATO ASI Ser. E, No. 100. Martinus Nijhoff, 1986, x + 308 pp, $44.50; $14.95 (P). [ISBN: 90-247-3234-4; 0-89838-850-3] 11 invited lectures from a NATO summer course in July 1985 in Corsica, France, together with various contributed seminar papers. Varied examples focus on patterns of random growth. How, for example, can random processes produce symmetric patterns? LAS

Applications (Physics), P. Lecture Notes in Physics-240: Monte-Carlo Methods and Applications in Neutronics, Photonics and Statistical Physics. Ed: R. Alcouffe, et al. Springer-Verlag, 1985, viii + 483 pp, $30 (P). [ISBN: 0-387-16070-1] Proceedings of an April 1985 joint U.S.-France conference on applications of Monte-Carlo methods to atomic physics. LAS

Applications (Social Science), S(15-17), P, L. Fair Allocation. Ed: H. Peyton Young. Proc. of Symp. in Appl. Math., V. 33. AMS, 1985, xiii + 170 pp, $26. [ISBN: 0-8218-0094-9] Six papers on aspects of fair division of power (apportionment, voting), money (income inequality, taxes, costs), and goods (auctions) from the January 1985 AMS short course at Anaheim, California. LAS

Reviewers

DFA: David F. Appleyard, Carleton; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; RD: Roger Day, St. Olaf; JD-B: John Dyer-Bennet, Carleton; GF: Giovanna Fjelstad, St. Olaf; SG: Steven Galovich, Carleton; BH: Bruce Hanson, St. Olaf; PH: Paul Humke, St. Olaf; KK: Kenneth Kaminsky, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; AM: Alan Magnuson, St. Olaf; SM: Steve McKelvey, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MT: Michael Tveite, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

We have solved the problem using only elementary algebra, but in solving an optimization problem it seems natural to employ the calculus. Now maximizing $(X/n)^n$ is equivalent to maximizing $n\ln(X/n)$ and so we consider the function $f(t) = t\ln(X/t)$. Since $f'(t) = \ln(X/t) - 1$, f has its maximum at $t = X/e$. In general this is not an integer and so the calculus is of no avail in solving this problem. However, in the special case when X/e is an integer n, then n is the unique solution and so must coincide with m in (3). So we have

$$n\left(1+\frac{1}{n-1}\right)^{n-1} < ne < (n+1)\left(1+\frac{1}{n}\right)^{n},$$

the inequality being strict since the solution is unique. Since this is true for all $n = 1, 2, \ldots$, we have inequality (1).

Reference

1. C. W. Barnes, Euler's constant and e, this MONTHLY, 91(1984) 428-430.

ON CHARACTERIZATIONS OF ANALYTICITY

G. B. FOLLAND

Department of Mathematics, University of Washington, Seattle, WA 98195

For the purposes of this note we adopt the following definition:

DEFINITION. Let U be an open set in the complex plane $\mathbb{C}$. A function $f: U \to \mathbb{C}$ is *analytic* if its complex derivative f' exists and is continuous on U. Equivalently, f is analytic if f is of class C^1 on U (as a function of two real variables) and satisfies the Cauchy-Riemann equation

$$\bar{\partial} f = 0, \quad \text{where } \bar{\partial} f = \frac{1}{2}\left(\frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y}\right). \tag{1}$$

The condition that f' be continuous is of course superfluous, and most complex analysis textbooks omit it in defining analyticity. However, there is a pedagogical advantage in the present definition: it enables one to prove Cauchy's theorem, and hence to penetrate to the heart of the subject, with considerably less effort. (For example, one can obtain a version of Cauchy's theorem sufficiently general for most purposes as an immediate corollary of Green's theorem.) Moreover, in the further development of function theory, the continuity of f' is never an issue; it always falls out of the calculations quite automatically.

At some point, however, one may wish to step back and ask whether the hypotheses of the theory can be weakened. The fact that f' need not be assumed continuous is the standard answer to this question. But another answer, more interesting in that it is an instance of a general result about solutions of elliptic differential equations (see [1, Corollary 6.31]), is that any *distribution* solution of the Cauchy-Riemann equation is an analytic function. We now give a simple version of this result which can be presented in a complex analysis class as soon as one has proved Weierstrass's theorem that the uniform limit of analytic functions is analytic.

NOTATION. If U is an open subset of $\mathbb{C}$, $C_0^1(U)$ denotes the space of functions of class C^1 on $\mathbb{C}$ (in the real-variable sense) that vanish outside a compact subset of U. dA denotes the element of area on $\mathbb{C}$. $\bar{\partial}$ denotes the Cauchy-Riemann operator defined in (1).

DEFINITION. A function $f: U \to \mathbb{C}$ is *weakly analytic* if f is continuous and

$$\int f \cdot \bar{\partial}\varphi \, dA = 0 \quad \text{for all } \varphi \in C_0^1(U). \tag{2}$$

It is clear that if f is analytic, then f is weakly analytic, for one can integrate by parts:

$$\int f \cdot \bar{\partial}\varphi \, dA = -\int \bar{\partial} f \cdot \varphi \, dA = 0.$$

(Note, however, that one needs some regularity of the first partial derivatives of f, for example, that f is of class C^1, for this argument to work: the mere differentiability of f would not suffice.) Conversely, we have:

THEOREM. *Every weakly analytic function is analytic.*

Proof. Suppose $f : U \to \mathbb{C}$ is weakly analytic. Choose a nonnegative function φ of class C^1 which vanishes outside the unit disc and satisfies $\int \varphi \, dA = 1$. (E.g., take

$$\varphi(z) = (3/\pi)(1 - |z|^2)^2 \quad \text{for } |z| < 1, \quad \varphi(z) = 0 \quad \text{for } |z| \geq 1.)$$

For $\delta > 0$, let $\varphi_\delta(z) = \delta^{-2}\varphi(z/\delta)$, and let U_δ be the set of points in U whose distance from the complement of U is greater than δ. If $z \in U_\delta$, the function $\psi(w) = \varphi_\delta(z - w)$ is clearly in $C_0^1(U)$, so the integral

$$f_\delta(z) = \int f(w)\varphi_\delta(z - w) \, dA(w)$$

is well defined. By differentiating under the integral, one sees that f_δ is of class C^1 on U_δ, and indeed (by (2)) that f_δ is analytic:

$$\bar{\partial} f_\delta(z) = \int f(w) \bar{\partial}\varphi_\delta(z - w) \, dA(w) = 0.$$

Since the U_δ's exhaust U as $\delta \to 0$, by Weierstrass's theorem it will suffice to show that $f_\delta \to f$ uniformly on K as $\delta \to 0$, for each compact $K \subset U$.

Given such a K, let η be the distance from K to the complement of U (if $U = \mathbb{C}$, let $\eta = 1$), and let $\tilde{K}$ be the set of points whose distance from K is at most $\eta/2$. $\tilde{K}$ is again a compact subset of U, so f is uniformly continuous on $\tilde{K}$. Hence, given $\varepsilon > 0$, we can choose $\delta \leq \eta/2$ small enough so that $|f(z) - f(w)| < \varepsilon$ whenever $|z - w| < \delta$ and $z, w \in \tilde{K}$. But then, since $\int \varphi_\delta dA = 1$ for all δ and $\varphi_\delta(z - w) = 0$ if $|z - w| \geq \delta$, for $z \in K$ we have

$$\begin{aligned} |f_\delta(z) - f(z)| &= \left| \int [f(w) - f(z)] \varphi_\delta(z - w) \, dA(w) \right| \\ &< \varepsilon \int \varphi_\delta(z - w) \, dA(w) \\ &= \varepsilon, \end{aligned}$$

and the proof is complete.

Reference

1. G. B. Folland, Introduction to Partial Differential Equations, Princeton University Press, Princeton, NJ, 1976.

SIMULTANEOUS COMPLEMENTS IN FINITE-DIMENSIONAL VECTOR SPACES

RICHARD D. BYRD
Department of Mathematics, University of Houston, Houston, TX 77004

Recall that if V is a vector space and U is a subspace of V, then a subspace X of V is called a complement to U in V if $V = U + X$ and $U \cap X = \{0\}$.

In [2], N. J. Lord proved the following theorem:

Let V be an n-dimensional vector space over the real numbers $\mathbb{R}$ (or the complex numbers), and

let $\{U_i: i \in \mathbb{N}\}$ be a countable family of subspaces of V each having the same dimension, k. Then there exists a subspace X of V which is a complement in V to all of the subspaces in the family.

To prove this theorem Lord introduces a topology, a family of continuous functions, and uses the Baire Category Theorem. Although the proof is neat, it has the disadvantage that most students in elementary linear algebra have never seen the Baire Category Theorem. (Many have no familarity with topological concepts.) We present below a proof of this theorem which would be accessible in elementary linear algebra. First we prove the following lemma, which is an obvious generalization of Problem 21, p. 177, of [**1**].

LEMMA. *Let V be an n-dimensional vector space over $\mathbb{R}$ and for each $i \in \mathbb{N}$, let U_i be a subspace of V such that* $\dim U_i < n$. *Then $V \neq \cup U_i$.*

Proof. If the lemma is not true, then there is a smallest positive integer n for which it is false. If the dimension of V is n, then there exists a countable collection of subspaces, U_i, of V such that $\dim U_i < n$ and $V = \cup U_i$. Since the dimension of V is finite, we may assume that under set inclusion, the members of this collection are pairwise incomparable. Thus for $i > 1$, $U_1 \cap U_i \neq U_1$, and by the minimality of n, there exists $u \in U_1$ with $u \notin \cup_{i=2}^{\infty}(U_1 \cap U_i)$. Pick v in V, not in U_1, and consider the set $T = \{v + au: a \in \mathbb{R}\}$. Since T is uncountable, there exists U_j such that $v + au$, $v + bu \in U_j$ and $a \neq b$. Consequently, $u \in U_j$ and so $j = 1$. But then $v \in U_1$, a contradiction.

Now for a proof of the theorem. If $k = n$, the zero subspace meets the condition. Thus we may assume that $k < n$. Let

$$\mathcal{T} = \{ m \in \mathbb{N}: \text{there exists an } m\text{-dimensional subspace } W \text{ of } V \text{ such that } W \cap U_i = \{0\} \text{ for each } i \in \mathbb{N}\}.$$

By the lemma, $1 \in \mathcal{T}$, and since V is finite-dimensional, $\mathcal{T}$ has a largest element, m. If X is an m-dimensional subspace of V with $X \cap U_i = \{0\}$ for each $i \in \mathbb{N}$, then it is easily verified that X is a complement in V to all of the subspaces in the family.

References

1. I. N. Herstein, Topics in Algebra, 2nd ed., Xerox College Pub., 1975.
2. N. J. Lord, Simultaneous complements in finite-dimensional vector spaces, this MONTHLY, 92(1985) 492–493.

OPPENHEIM'S INEQUALITY FOR POSITIVE DEFINITE MATRICES

THOMAS L. MARKHAM

Department of Mathematics, University of South Carolina, Columbia, SC 29208

In 1930, Sir Alexander Oppenheim [7] proved a beautiful determinantal inequality for positive definite matrices. Various analogs of this result have been realized for other classes of matrices [**1**], [**3**], [**4**], [**5**].

We plan to offer a new proof of Oppenheim's result that is straightforward and tends to illuminate the case of equality.

First, we introduce some useful notation. If A is an $n \times n$ matrix over the complex field, consider the partition of A

$$(1) \qquad \begin{pmatrix} a_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad \text{where } A_{22} \text{ has order } n-1.$$

Now suppose $a_{11} \neq 0$. The Schur complement [**2**] of a_{11} in A, denoted $A|a_{11}$, is defined by

$$A|a_{11} := A_{22} - A_{21}a_{11}^{-1}A_{12}.$$

If $A = (a_{ij})$ and $B = (b_{ij})$ are $n \times n$ matrices, the Hadamard product of A and B is the $n \times n$ matrix $A \circ B := (a_{ij} b_{ij})$.

Now we state and prove Oppenheim's result.

THEOREM. *If A and B are $n \times n$ positive definite matrices, then* $\det(A \circ B) \geq (\Pi_{i=1}^{n} a_{ii})\det(B)$, *and equality holds if and only if B is a diagonal matrix.*

Proof.

$$\text{Partition } B = \begin{pmatrix} b_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}, \quad \text{where } B_{22} \text{ is of order } n-1,$$

and let $E = \dfrac{1}{b_{11}} B_{21} B_{12}$. We will use induction on the order of A and B. Since $n = 1$ is obvious, we assume $n > 1$ and the result holds for positive definite matrices of order $K < n$.

$$\begin{aligned} \text{Now } \det(A \circ B) &= a_{11} b_{11} \det(A \circ B | (a_{11} b_{11})) \\ &= a_{11} b_{11} \det[A_{22} \circ (B | b_{11}) + (A | a_{11}) \circ E]. \end{aligned}$$

It is well known that if B is positive definite, then $B|b_{11}$ is positive definite. Also, Schur [**8**] proved that the Hadamard product of two positive definite matrices is again positive definite. Thus $A_{22} \circ (B|b_{11})$ is positive definite, and $(A|a_{11}) \circ E$ is positive definite if E has a positive diagonal. Minkowski's inequality [**6**, pp. 419–420] allows us to say

$$\det[A_{22} \circ (B|b_{11}) + A|a_{11} \circ E] \geq \det A_{22} \circ (B|b_{11}) + \det(A|a_{11}) \circ E. \tag{2}$$

Using our inductive hypothesis and the fact that $\det(A|a_{11} \circ E) \geq 0$, we have

$$\begin{aligned} \det(A \circ B) &\geq a_{11} b_{11} \det(A_{22} \circ B|b_{11}) \\ &\geq a_{11} b_{11} [a_{22} \cdots a_{nn} \det(B|b_{11})] \\ &= \left(\prod_{i=1}^{n} a_{ii}\right) \det B. \end{aligned}$$

Next, suppose that equality holds; i.e., $\det(A \circ B) = a_{11} \cdots a_{nn} \det B$. Then we must have

$$\det(A \circ B | a_{11} b_{11}) = \det(A_{22} \circ B|b_{11}) + \det(A|a_{11} \circ E).$$

This can happen in this case only if $A|a_{11} \circ E = 0$. Thus, $b_{i1} b_{1i} = 0$ for all $i = 2, \ldots, n$. Hence $b_{i1} = 0$ for $i = 2, \ldots, n$, since B is hermitian.

Since the result is invariant under simultaneous permutational similarity of A and B, it is clear that any row of B could be permuted to the first row, where off-diagonal entries of that row are permuted to off-diagonal entries of the first row. Thus, all off-diagonal entries of B are zero, and B is a diagonal matrix.

If B is diagonal, then it is obvious that

$$\det(A \circ B) = \left(\prod_{i=1}^{n} a_{ii}\right) \det B. \quad \blacksquare$$

References

1. M. Fiedler and V. Pták, Diagonally dominant matrices, Czechoslovak Math. J., 17(92) (1967) 420–433.

2. E. Haynsworth, Determination of the inertia of a partitioned Hermitian matrix, Linear Algebra and Appl., 1 (1968) 73–81.

3. M. S. Lynn, On the Schur product of the H-matrices and nonnegative matrices, and related inequalities, Proc. Cambridge Phil. Soc., 60 (1964) 425–431.

4. T. L. Markham, A semigroup of totally nonnegative matrices, Linear Algebra and Appl., 3 (1970) 157–164.

5. ______, Class G, and a unification problem, Linear Algebra and Appl., 39 (1981) 197–204.

6. L. Mirsky, An Introduction to Linear Algebra, Oxford University Press, Oxford, 1955.

7. A. Oppenheim, Inequalities connected with definite Hermitian forms, J. London Math. Soc., 5 (1930) 114–119.

8. I. Schur, Bemerkungen Zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen, J. Reine Math., 140 (1911) 1–28.

THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

For instructions about submitting material for publication in this department see the inside front cover.

COUNTEREXAMPLES TO L'HÔPITAL'S RULE

R. P. BOAS

Department of Mathematics, Northwestern University, Evanston, IL 60201

1. Introduction. I am not, of course, claiming that L'Hôpital's rule is wrong, merely that unless it is both stated and used very carefully it is capable of yielding spurious results. This is not a new observation, but it is often overlooked.

For definiteness, let us consider the version of the rule that says that if f and g are differentiable in an interval (a, b), if

$$\lim_{x\to b-} f(x) = \lim_{x\to b-} g(x) = \infty,$$

and if $g'(x) \neq 0$ *in some interval* (c, b), then

$$\lim_{x\to b-} f'(x)/g'(x) = L$$

implies that

$$\lim_{x\to b-} f(x)/g(x) = L.$$

If $\lim f'(x)/g'(x)$ does not exist, we are not entitled to draw any conclusion about $\lim f(x)/g(x)$. Strictly speaking, if g' has zeros in every left-hand neighborhood of b, then f'/g' is not defined on (a, b), and we ought to say firmly that $\lim f'/g'$ does not exist. There is, however, the insidious possibility that f' and g' contain a common factor: $f'(x) = s(x)\psi(x)$, $g'(x) = s(x)\omega(x)$, where s does not approach a limit and $\lim \psi(x)/\omega(x)$ exists. It is then quite natural to cancel the factor $s(x)$. This is just what we must not do in the present situation: it is quite possible that $\lim \psi(x)/\omega(x)$ exists but $\lim f(x)/g(x)$ does not.

This claim calls for an example. A number of textbooks give one, but it is (as far as I know) always the same example. The aim of this note is both to emphasize the necessity of the condition $g'(x) \neq 0$ and to provide a systematic method of constructing counterexamples when this condition is violated. I consider the case when $b = +\infty$, since the formulas are simpler than when b is finite.

2. A construction. Take a periodic function λ (not a constant) with a bounded derivative, for example $\lambda(x) = \sin x$. Let

$$f(x) = \int_0^x \{\lambda'(t)\}^2\, dt.$$

It is clear that $f(x) \to +\infty$ as $x \to +\infty$. Now choose a function φ such that $\varphi(\lambda(x))$ is bounded and both $\varphi(\lambda(x))$ and $\varphi'(\lambda(x))$ are bounded away from 0. There are many such

functions φ; for example,

$$\varphi(x)=e^{x}\quad\text{or}\quad(x+c)^2\quad\text{or}\quad 1/(c+x),$$

provided $|\lambda(x)|<c$ and $|\lambda'(x)|<c$. Take $g(x)$ to be $f(x)\varphi(\lambda(x))$. Since $\inf\varphi(\lambda(x))>0$, we have $g(x)\to\infty$ as $x\to\infty$.

Now try to apply L'Hôpital's rule to $f(x)/g(x)$. We have to consider $f'(x)/g'(x)$, where

$$f'(x)=\{\lambda'(x)\}^2,$$
$$g'(x)=\{\lambda'(x)\}^2\varphi(\lambda(x))+f(x)\varphi'(\lambda(x))\lambda'(x).$$

Here $g'(x)=0$ whenever $\lambda'(x)=0$, i.e., g' has zeros in every neighborhood of ∞, and consequently we are not entitled to apply L'Hôpital's rule at all. However, this conclusion seems rather pedantic; let us go ahead anyway. If we cancel the factor $\lambda'(x)$, we obtain

$$\frac{f'(x)}{g'(x)}=\frac{\lambda'(x)}{\lambda'(x)\varphi(\lambda(x))+f(x)\varphi'(\lambda(x))}.$$

Now $\lambda'(x)$ is bounded (by hypothesis), $\lambda'(x)\varphi(\lambda(x))$ is bounded, $\varphi'(\lambda(x))$ is bounded away from 0, but $f(x)\to\infty$, so $f'(x)/g'(x)\to 0$. Yet $f(x)/g(x)=1/\varphi(\lambda(x))$ does not approach zero, since $\varphi(\lambda(x))$ is bounded!

3. Discussion. What went wrong? If you will study any proof of L'Hôpital's rule, you will find a place where it used (or should have used) the assumption that $g'(x)$ did not change sign infinitely often in a neighborhood of ∞. Our example shows that, at least sometimes, L'Hôpital's rule actually fails when this hypothesis is not satisfied.

The phenomenon just described was discovered more than a century ago by O. Stolz [**1**], [**2**]. His example was $\lambda(x)=\sin x$, $\varphi(x)=e^{x}$; it has been repeated in all the modern discussions that I have seen. It was wondering whether there *are* any other examples that led to this note.

One can verify that it is the changes of sign of $\lambda'(x)$ that cause the trouble, not the mere presence of zeros of λ'. In other words, if $\lambda'(x)\geqslant 0$, the cancellation process still leads to a correct result, as Stolz pointed out. However, it seems wildly improbable that an example of either kind will occur in practice, especially for limits at a finite point. Differentiable functions with infinitely many changes of sign in a finite interval are rarely encountered outside notes like this one; all the less, functions with infinitely many double zeros.

4. History. Guillaume François Antoine de Lhospital, Marquis de Sainte-Mesme (1651–1704) published (anonymously) in 1691 the world's first textbook on calculus, based on John Bernoulli's lecture notes. He seems to have written his name as above, but it is more familiar as L'Hospital (old French spelling) or L'Hôpital (modern French); I prefer the latter, since it stops students from pronouncing the s (which Larousse's dictionary says is not to be pronounced).

References

1. O. Stolz, Ueber die Grenzwerthe der Quotienten, Math. Ann., 15 (1879) 556–559.
2. ______, Grundzüge der Differential- und Integralrechnung, vol. 1, Teubner, Leipzig, 1893, pp. 72–84.

CONVOLUTIONS OF CAUCHY DISTRIBUTIONS

Colin R. Blyth
Department of Mathematics and Statistics, Queen's University, Kingston, Ontario, Canada K7L 3N6

Recently in this Monthly Dwass [**1**] and Nelson [**3**] have discussed finding the distribution of a sum of two independent Cauchy random variables using the convolution formula with partial

fraction decomposition of its integrand. Lehmann [2] gives this as a problem for two independent Cauchy (a_i, b_i), density $(b_i/\pi)/\{b_i^2 + (x - a_i)^2\}$ with $b_i > 0$, random variables X_i $(i = 1,2)$.

Simplifying transformations are the key to solving problems like this one efficiently. It is much easier to work with

$$X_1' = \frac{X_1 - a_1}{b_1} \quad \text{and} \quad X_2' = \frac{X_2 - a_2}{b_2}$$

which have independent standard Cauchy $(0,1)$ distributions. And instead of finding the density of

$$X_1 + X_2 = b_1 X_1' + b_2 X_2' + (a_1 + a_2) \tag{1}$$

it is easier to first find the density of

$$\begin{aligned} V &= \frac{X_1 + X_2 - (a_1 + a_2)}{b_2} = \frac{b_1}{b_2} X_1' + X_2' \\ &= kX_1' + X_2', \quad \text{writing } k \text{ for } b_1/b_2. \end{aligned}$$

The convolution formula for the density of V at v is

$$g(v) = \frac{1}{\pi^2} \int_{-\infty}^{\infty} \frac{1}{1 + x^2} \cdot \frac{1}{1 + (kx - v)^2}\, dx. \tag{2}$$

Integrating this is a standard problem in elementary calculus:

$$g(v) = \frac{1}{\pi^2} \int_{-\infty}^{\infty} \left\{ \frac{A + Bx}{1 + x^2} + \frac{kC + kD(kx - v)}{1 + (kx - v)^2} \right\} dx. \tag{3}$$

This integral can be written down, so the only problem is to find the constants A, B, C, D that make (3) coincide with (2). Brought to a common denominator, the integrand in (3) has for its numerator a cubic in x whose constant term must be 1 and whose other coefficients must all be 0. From the coefficient of v^3 we have at once $D = -B$, giving

$$\begin{aligned} g(v) &= \frac{1}{\pi^2} \left[A \tan^{-1} x + C \tan^{-1}(kx - v) + \frac{B}{2} \log \frac{1 + x^2}{1 + (kx - v)^2} \right]_{-\infty}^{\infty} \\ &= \frac{1}{\pi}(A + C). \end{aligned} \tag{4}$$

The other three equations, for A, B, C, are

$$(1 + v^2)A + kvB + kC = 1, \tag{5}$$

$$-2kvA + (1 - k^2 + v^2)B = 0, \tag{6}$$

$$kA - vB + C = 0. \tag{7}$$

Since only $A + C$ is needed, we can save work by not fully solving these equations. Use $vB = kA + C$ from (7) to eliminate B from (5) and from (6) multiplied through by v:

$$(1 + k^2 + v^2)A + 2kC = 1, \tag{8}$$

$$k(1 - k^2 - v^2)A + (1 - k^2 + v^2)C = 0. \tag{9}$$

Multiply (8) by $(1 + k)$ and add to (9):

$$(1 + 2k + k^2 + v^2)A + (1 + 2k + k^2 + v^2)C = 1 + k.$$

Therefore

$$A + C = \frac{1 + k}{(1 + k)^2 + v^2},$$

showing in (4) that V has Cauchy $(0, 1 + k)$ distribution, and it follows at once that

$$X_1 + X_2 = b_2 V + (a_1 + a_2)$$

has Cauchy $(a_1 + a_2, b_2(1 + k))$ = Cauchy $(a_1 + a_2, b_1 + b_2)$ distribution.

NOTE 1. For $a_1 = a_2 = 0$ and $b_1 + b_2 = 1$, the distribution of $X_1 + X_2 = b_1 X_1' + b_2 X_2'$ is Cauchy $(0, 1)$: Every weighted average (recall, $b_1 > 0$ and $b_2 > 0$) of two independent Cauchy $(0, 1)$ random variables has Cauchy $(0, 1)$ distribution.

NOTE 2. The apparently more general problem of finding the distribution of $c_1 X_1 + c_2 X_2 + d$, with each c_i possibly negative, is also covered, because

$$c_1 X_1 + c_2 X_2 + d = b_1 c_1 X_1' + b_2 c_2 X_2' + a_1 c_1 + a_2 c_2 + d$$

has the same distribution as

$$|b_1 c_1| X_1' + |b_2 c_2| X_2' + a_1 c_1 + a_2 c_2 + d$$

since X_i' and $-X_i'$ have the same distribution. This last expression has the form (1).

NOTE 3. If U has Uniform $(-\pi/2, \pi/2)$ distribution, then $X = \tan U$ has Cauchy $(0, 1)$ distribution. There ought to be an almost instant convolution proof, working entirely in terms of Uniform distributions—perhaps geometrically, using circles tangent to a common X-axis. Is there?

This work was supported by the Natural Sciences and Engineering Research Council of Canada.

References

1. Meyer Dwass, On the convolution of Cauchy distributions, this MONTHLY, 92 (1985) 55–57.
2. Erich L. Lehmann, Theory of Point Estimation, Wiley, New York, 1983, p. 58.
3. Roger B. Nelsen, Letter to the Editor, this MONTHLY, 92 (1985) 679.

A METHOD FOR FINDING THE EIGENVECTORS OF AN $n \times n$ MATRIX CORRESPONDING TO EIGENVALUES OF MULTIPLICITY ONE

M. CARCHIDI
Department of Physics and Atmospheric Science, Drexel University, Philadelphia, PA 19104

Many texts [1], [2] and introductory courses on linear algebra introduce the adjoint matrix (the transpose of the cofactor matrix) to determine an expression for the inverse of an $n \times n$ invertible matrix. The purpose of this report is to show, using elementary linear algebra techniques, that the adjoint matrix can also be used to find the eigenvectors of an $n \times n$ matrix corresponding to eigenvalues of multiplicity one. An advanced analysis of our results lying outside the scope of an introductory linear algebra course can be found in the text *Matrix Theory* by F. R. Gantmacher [3].

As mentioned, the introductory linear algebra student's first encounter with the adjoint matrix appears in the expression [1], [2]

$$A[adj(A)] = (\det A) I, \tag{1}$$

where $adj(A)$ is the adjoint matrix of an $n \times n$ matrix A, and $\det A$ is its determinant. When $\det A \neq 0$, Equation (1) leads to the usual formula,

$$A^{-1} = [\det A]^{-1}[adj(A)]$$

for the inverse of the matrix. Equation (1), however, is usually not explored in an introductory course for the case when $\det A = 0$. We will consider one implication of this case. When $\det A = 0$, Equation (1) becomes

$$A[adj(A)] = 0. \tag{2}$$

If λ_i is an eigenvalue of a matrix H (making $\det(\lambda_i I - H) = 0$), then the substitution $A = \lambda_i I - H$ into Equation (2) gives $(\lambda_i I - H)[adj(\lambda_i I - H)] = 0$ or

$$H[adj(\lambda_i I - H)] = \lambda_i[adj(\lambda_i I - H)]. \tag{3}$$

The pth column of this matrix equation is

$$H[adj(\lambda_i I - H)]_p = \lambda_i[adj(\lambda_i I - H)]_p, \tag{4}$$

which shows that any nonzero column $[adj(\lambda_i I - H)]_p$ of $adj(\lambda_i I - H)$ is an eigenvector for H corresponding to the eigenvalue λ_i.

As an example of this method consider the 3×3 matrix

$$H = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 4 & -17 & 8 \end{bmatrix}$$

which has $\lambda_i = 4$ as an eigenvalue of multiplicity one. Then

$$4I - H = \begin{bmatrix} 4 & -1 & 0 \\ 0 & 4 & -1 \\ -4 & 17 & -4 \end{bmatrix} \quad \text{and} \quad adj(4I - H) = \begin{bmatrix} 1 & -4 & 1 \\ 4 & -16 & 4 \\ 16 & -64 & 16 \end{bmatrix}.$$

Since $[adj(4I - H)]_{11}$ is not zero, we can take $p = 1$, giving $[1\ 4\ 16]^{\mathrm{T}}$ as an eigenvector for H corresponding to the eigenvalue $\lambda_i = 4$. It is usually not necessary to calculate the entire $adj(\lambda_i I - H)$ matrix. In the example above, once we find that $[adj(4I - H)]_{11}$ is not zero, then we need only find the 1st column of $adj(\lambda_i I - H)$ when obtaining the eigenvector. In fact, for a 2×2 or 3×3 matrix this technique is much faster and easier to program on a computer than the usual methods of computing eigenvectors.

The only complication to the method above arises when $adj(\lambda_i I - H)$ is the zero matrix. To address this complication, we use the following result from linear algebra for the derivative of $\det(\lambda I - H)$ [4]:

$$(d/d\lambda)[\det(\lambda I - H)] = \operatorname{Trace}\{adj(\lambda I - H)\}. \tag{5}$$

We shall apply this result to the eigenvalue polynomial

$$\det(\lambda I - H) = \prod_{j=1}^{n} (\lambda - \lambda_j),$$

where $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_n$ are the eigenvalues of H. Taking the derivative of both sides of this equation with respect to λ and using Equation (5) leads to

$$\operatorname{Trace}\{adj(\lambda I - H)\} = \sum_{k=1}^{n} \prod_{\substack{j=1 \\ j \neq k}}^{n} (\lambda - \lambda_j). \tag{6}$$

By setting $\lambda = \lambda_i$ in Equation (6) we obtain

$$\operatorname{Trace}\{adj(\lambda_i I - H)\} = \sum_{j=1}^{n} [adj(\lambda_i I - H)]_{jj} = \prod_{\substack{j=1 \\ j \neq i}}^{n} (\lambda_i - \lambda_j). \tag{7}$$

Therefore if λ_i is an eigenvalue of multiplicity one, the righthand side of Equation (7) is nonzero. It follows that there exists a value of p between 1 and n for which $[adj(\lambda_i I - H)]_{pp}$ is nonzero, and hence $adj(\lambda_i I - H) \neq 0$.

To summarize: Let H be an $n \times n$ matrix with eigenvalues $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_n$. Further, let λ_i for some i have multiplicity one. Then there exists a value of p such that the pth column of $adj(\lambda_i I - H)$ is an eigenvector for H corresponding to λ_i. Moreover, a value of p to use is one for which $[adj(\lambda_i I - H)]_{pp}$ is not zero (i.e., we need only consider the cofactors of the diagonal elements of $\lambda_i I - H$ when determining p). Of course, once one eigenvector corresponding to an eigenvalue of multiplicity one is known, all others are just multiples of it.

For the case when the multiplicity of λ_i is greater than one, it can be shown [3] that the matrix $adj(\lambda_i I - H)$ is nonzero provided the eigenspace of eigenvectors corresponding to the eigenvalue λ_i has dimension exactly equal to one. However, in this case $adj[(\lambda_i I - H)]_{pp}$ may be zero for all p, and so the choice of a suitable value of p may be more difficult.

Acknowledgement. I wish to thank Dr. C. Rorres for reviewing this manuscript.

References

1. Howard Anton, Elementary Linear Algebra, 4th ed., Wiley, New York, 1984.
2. Bernard Kolman, Introductory Linear Algebra with Applications, 3rd ed., Macmillan, New York, 1984.
3. F. R. Gantmacher, Matrix Theory, Vol. I, Chelsea, New York, 1960.
4. Charles G. Cullen, Matrices and Linear Transformations (Theorem 8–15, page 183), Addison-Wesley, Reading, MA, 1966.

PROBLEMS AND SOLUTIONS

*For instructions about submitting **proposed** problems for publication in this department see the inside front cover. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the address given on the inside front cover.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

For instructions about submitting solutions of these Elementary Problems, which should be mailed by February 28, 1987, see the inside front cover. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgement).

E 3165. *Proposed by Liviu I. Nicolaescu (student), Iassy, Romania.*

Let R be a finite commutative ring with 1, $|R| = k$. Let $N = N(R)$ denote the set of nonzero non-units of R. Prove that $N \neq \varnothing$ forces $|N| \geqslant [\sqrt{k-1}\,]$, where $[x]$ denotes the greatest integer in the real number x.

E 3166. *Proposed by Armel Mercier, University of Quebec.*

Let $n \geqslant 1$ be an integer and let f be a polynomial of degree less than or equal to n. Show the following combinatorial identity:

$$\sum_{j=1}^{n} \frac{(-1)^{j+1}\binom{n}{j} f(2j+1)}{2j(2j+1)} = \frac{n!(n+1)!2^{2n+1} f(0)}{(2n+2)!} + f'(1) - f(1) + f(1) \sum_{j=1}^{n} \frac{1}{2j}.$$

E 3167. *Proposed by Ernesto Bruno Cossi, Universidade Federal Do Rio Grande Do Sul, Brazil.*

Let P_i be any one of the five regular polyhedra inscribed in a unit sphere. For each polyhedron P_i, determine the smallest and the largest number of vertices of P_i which can be seen from a point on a concentric sphere of radius $R > 1$.

E 3168. *Proposed by Douglas B. Tyler, University of California, Davis.*

(i) Show that if

$$J(a, n) = \int_0^{\infty} [x(x+a)]^{(n/2)-1} \left[\ln\left(\frac{x}{x+a}\right)\right]^n dx,$$

then $J(a, n) = (-1)^n a^{n-1} b_n$, where

$$b_n = \int_{-\infty}^{\infty} \left(\frac{y}{\sinh(y)}\right)^n dy.$$

(ii) Show that b_n is a polynomial of degree $[(n+1)/2]$ in π^2 with rational coefficients for $n = 1, 2, 3, \ldots$. (See E 2865 [1981, 66]; [1982, 426].)

*(iii) Find a closed form expression for the coefficients.

E 3169. *Proposed by Dean S. Clark, University of Rhode Island.*

A gentleman, who decides randomly what to eat for dinner so long as it isn't the same thing two nights in a row, will return to his favorite restaurant next Friday. Is it more likely that he will order (a) Chicken Mediterranean or (b) the dinner he had (which he cannot remember) when he first visited the restaurant? How much more likely?

E 3170. *Proposed by The Howard University Group, Washington, DC.*

Construct a graph as follows: Put $n + 1$ labeled vertices around a circle and let the edges be the straight line segments connecting any two vertices. A tree is non-crossing if no two edges intersect except at the vertices. Enumerate the number of non-crossing spanning trees for this graph. For $n = 1, 2, 3$, the numbers are 1, 3, 12, respectively.

SOLUTIONS OF ELEMENTARY PROBLEMS

An "Equal Perimeters" Point for Three Hyperbolas

E 3020 [1983, 644]. *Proposed by Clark Kimberling, University of Evansville.*

Suppose ABC is a nonisosceles triangle. Find three hyperbolas concurrent in a point P such that triangles APB, APC, and BPC all have the same perimeter. (*) How does this common perimeter compare with that of ABC?

Solution by R. W. Wagner, University of Massachusetts, Amherst. Let the lengths of the sides of

the original nonisosceles triangle, AB, AC, and BC be $2c, 2b, 2$, respectively, with $c < b < 1$, after relabelling the vertices if necessary.

Equality of perimeters requires that

$$PA + PB + 2c = PA + PC + 2b = PB + PC + 2.$$

From the first pair, $PB - PC = 2(b - c)$. This requires that P be on one branch of the hyperbola with foci at B and C and with vertices a distance $(b - c)$ from the midpoint of BC. Moreover, the branch is the one nearer the focus opposite the shorter side. Similarly for the branches of the other two hyperbolas which will have the other pairs of vertices for foci.

To examine the ratio of the common perimeter to that of the original triangle, $2b + 2c + 2$, introduce coordinates so that B: $(-1, 0)$, C: $(1, 0)$, and P: (x, y). Let A: (r, s) with $s > 0$. To insure that P is on the correct branch of the first hyperbola mentioned, use the parametric equations for this hyperbola and set

$$x = (b - c)\sec\theta > 0, \quad y = \sqrt{1 - (b - c)^2}\tan\theta.$$

A standard property of this parametrization is that

$$PB = \sec\theta + (b - c) \quad \text{and} \quad PC = \sec\theta - (b - c),$$

so that the perimeter of triangle PBC is $2 + 2\sec\theta$. The ratio to be studied is

$$\frac{2 + 2\sec\theta}{2 + 2b + 2c} = \frac{1 + \sec\theta}{1 + b + c}.$$

The point P must also lie on the branch of the hyperbola with foci A and B and distance between vertices $2(1 - b)$ nearer to B. This requires that

$$PA = PB + 2(1 - b) = \sec\theta + 2 - (b + c).$$

Using the coordinates r and s of A, we see this implies that

$$[(b - c)\sec\theta - r]^2 + \left[\sqrt{1 - (b - c)^2}\tan\theta - s\right]^2 = [\sec\theta + 2 - (b + c)]^2.$$

When this is expanded, the identity $\tan^2\theta = \sec^2\theta - 1$ can be used to eliminate these second degree terms and the other terms can be combined to get an equation of the form

$$L\sec\theta - M\tan\theta + N = 0,$$

wherein

$$\begin{aligned} L &= -2(b - c)r - 2[2 - (b + c)], \\ M &= -2s\sqrt{1 - (b - c)^2}, \\ N &= r^2 + s^2 - \left[1 - (b - c)^2\right] - [2 - (b + c)]^2. \end{aligned}$$

However, r and s are related to b and c by

$$(r - 1)^2 + s^2 = 4b^2 \quad \text{and} \quad (r + 1)^2 + s^2 = 4c^2.$$

Subtracting the first from the second leads to

$$r = c^2 - b^2$$

which is negative, and adding leads to $r^2 + s^2 = 2(b^2 + c^2) - 1$, and after recognizing that $2(b^2 + c^2) = (b + c)^2 + (b - c)^2$ to

$$s^2 = \left[1 - (b - c)^2\right]\left[(b + c)^2 - 1\right].$$

Note that both factors are positive; the second because of the triangle inequality. Finally, in terms of b and c only, after dropping a common factor,

$$L = (b-c)^2(b+c) + (b+c) - 2,$$
$$M = \left[1-(b-c)^2\right]\sqrt{(b+c)^2 - 1},$$
$$N = (b-c)^2 + 2(b+c) - 3.$$

To avoid more extraneous roots change the parameter by

$$\sec\theta = \frac{1+t^2}{1-t^2}, \quad \tan\theta = \frac{2t}{1-t^2}.$$

Then the equation to solve for t is

$$f(t) = (N-L)t^2 + 2Mt - (N+L) = 0$$

and the roots are

$$t = \frac{-M \pm \sqrt{M^2+N^2-L^2}}{N-L}.$$

One can verify that

$$N - L = (b+c-1)\left[1-(b-c)^2\right] > 0,$$
$$M^2 + N^2 - L^2 = 4\left[1-(b-c)^2\right](b+c-1)^2 > 0,$$

that the minimum value of $f(t)$ is negative and occurs at

$$t = -M/(N-L) = -\sqrt{(b+c+1)/(b+c-1)} < -1.$$

In order for P to be on the correct branch of the hyperbola, $f(t)$ must have a zero greater than -1. Thus, to get an equal perimeters point, one must have

$$f(-1) = 2(M-L) < 0.$$

The graph of $M = L$ in the b, c-plane is a curve with a fourth degree polynomial equation and meets the boundary of the triangular domain of b and c, $b < c < 1 - b$, $\frac{1}{2} < b < 1$, at $(5/8, 5/8)$ and at $(1, 0)$. In order to have an equal perimeters point the point determining the triangle must be above this curve.

Because $1 + \sec\theta = 2/(1-t^2)$, the ratio of perimeters is $2/(1+b+c)(1-t^2)$. The values of this ratio have the value $(3+2\sqrt{3})/9$, associated with an equilateral triangle, for a greatest lower bound but no upper bound. The reason for this is that near $L = M$, the value of t is near -1.

Also solved in entirety by J. Dou (Spain) and P. Yff. The first part of the problem was solved by L. Kuipers (Switzerland) and the proposer.

For a different and more general approach to the solution of this problem, see the article titled *The isoperimetric point and the point(s) of equal detour in a triangle* by G. R. Veldkamp in this MONTHLY [1985, 546–58].

A Hypergeometric Polynomial

E 3021 [1983, 645]. *Proposed by Geng-Zhe Chang and Edward T. H. Wang, Wilfrid Laurier University.*

Let

$$p_n(x) = \sum_{k=0}^{n} \binom{n}{k}^2 (1+x)^k (1-x)^{n-k}.$$

Express $p_n(x)$ as an explicit function of $1 - x^2$.

Solution I *by O. P. Lossers, Eindhoven University of Technology, The Netherlands.* It is well known (see [**1**]) that

$$\sum_{k=0}^{n}\left[\frac{(-n)_k}{k!}\right]^2 y^k = \sum_{k=0}^{n}\binom{n}{k}^2 y^k = (1-y)^n P_n\left(\frac{1+y}{1-y}\right),$$

where P_n is the nth Legendre polynomial. So

$$\sum_{k=0}^{n}\binom{n}{k}^2(1+x)^k(1-x)^{n-k} = (1-x)^n\left[1-\frac{1+x}{1-x}\right]^n P_n\left(-\frac{1}{x}\right)$$
$$= (-2x)^n P_n\left(-\frac{1}{x}\right) = (2x)^n P_n\left(+\frac{1}{x}\right).$$

Furthermore (see [**2**])

$$P_n\left(\frac{1}{x}\right) = x^{-n}{}_2F_1\left(-\frac{1}{2}n, \frac{1}{2}-\frac{1}{2}n, 1, 1-x^2\right).$$

So

$$p_n(x) = 2^n{}_2F_1\left(-\frac{n}{2}, \frac{1-n}{2}, 1, 1-x^2\right),$$

and this hypergeometric function indeed is a polynomial in the variable $1 - x^2$.

References

1. E. R. Hansen, A Table of Series and Products, Prentice-Hall, 1975, formula 10.24.2.
2. A. Erdélyi et al., Higher Transcendental Functions, vol. I, McGraw-Hill, 1953, p. 129, formula 2.4.

Solution II *by Roger B. Nelsen, Lewis and Clark College, Portland, Oregon.* Since $p_n(x)$ can be written as

$$p_n(x) = \sum_{k=0}^{n}\binom{n}{k}(1+x)^k\cdot\binom{n}{n-k}(1-x)^{n-k},$$

it represents the coefficient of θ^n in the expansion of

$$f(\theta) = [(1+x)+\theta]^n[(1-x)+\theta]^n,$$

obtained by multiplying the binomial expansions of each factor.

However, we also have $f(\theta) = [(1-x^2) + 2\theta + \theta^2]^n$, which has the expansion

$$f(\theta) = \sum_{\substack{i,j,k\geqslant 0\\ i+j+k=n}}\frac{n!}{i!j!k!}(1-x^2)^i(2\theta)^j(\theta^2)^k.$$

Terms in θ^n occur for $j + 2k = n$, or $i = k$. Hence we have

$$p_n(x) = \sum_{k=0}^{[\![\frac{n}{2}]\!]}\frac{n!}{k!k!(n-2k)!}2^{n-2k}(1-x^2)^k,$$

where $[\![\,\cdot\,]\!]$ represents the greatest integer function.

Also solved by 37 other readers and the proposers. Robert Shafer pointed out a similarity to problem 81-13 (SIAM REVIEW, July 1982). Richard Askey observes that this is a special case of an 1836 formula of Kummer. Pei Yuan Wu points out that any even polynomial can be expressed as a polynomial in $1 - x^2$.

Prime Divisor of a Sum

E 3024 [1983, 645]. *Proposed by Eugene M. Luks, University of Oregon.*

Let a, b be integers with $b > 0$. Prove that there are infinitely many positive integers n with the property that if p is a prime divisor of $n^b + a$, then p is also a divisor of $k^b + a$ for some integer k with $|k|^b < n$. (The case $b = 2$, $a = 3$ was problem $B - 3$ on the 1981 William Lowell Putnam Competition.)

Solution by the proposer. We use the following

LEMMA. *Let r be a positive integer. There exist positive integers s, t_0 such that if t is any integer larger than t_0, then the prime divisors of $t^s + 1$ and $t^s - 1$ are less than $t^{s/r}$.*

Proof of lemma . Let s be a product of distinct odd primes, not including any prime divisors of r, such that

$$\frac{\phi(s)}{s} < \frac{1}{r}.$$

Such a product exists for

$$\prod_{p \text{ prime}} \frac{p-1}{p} = 0.$$

Now suppose $f(x)$ is any monic irreducible factor of $x^s - 1$ in $\mathbf{Z}[x]$. The roots of $f(x)$ are the primitive qth roots of unity for some $q|s$, so

$$\deg(f) = \phi(q) \leqslant \phi(s) < \frac{s}{r}.$$

Hence, there is some positive t_0 such that $t > t_0$ implies $|f(t)| < t^{s/r}$ and $|f(-t)| < t^{s/r}$ for all such factors f. Since any prime divisor of $t^s - 1$ divides some $f(t)$ and any prime divisor of $t^s + 1 = -((-t)^s - 1)$ divides some $f(-t)$, we are done. □

Proof of problem statement. Let $r = b^2$ and pick s, t_0 according to the lemma.

CASE 1. $a = 0$.

Let $n = 2^c$ for any $c > b$. If p is prime and $p|n^b$, then $p = 2$. Thus, since $p|2^b$ and $2^b < n$, we may choose $k = 2$.

CASE 2. $|a| = 1$.

Take any m with $m^b > t_0$ and let $n = m^s$. By the lemma, if p is prime and p divides $n^b + a = m^{sb} \pm 1$, then $p^r < m^{sb} = n^b$ so that $p^b < n$. Thus, a suitable k is given by $k \equiv n \bmod(p)$, $0 < k < p$.

CASE 3. $|a| > 1$.

Since $(b, s) = 1$, there exist positive integers u, v satisfying $ub = vs + 1$. There is an m_0 such that $m \geqslant m_0$ implies

$$\text{(i) } u + ms > b \quad \text{and} \quad \text{(ii) } |a|^{v+mb} > t_0.$$

Then for each $m \geqslant m_0$, consider

$$n = |a|^{u+ms}$$

so that

$$n^b + a = \pm a(a^{vs+mbs} \pm 1).$$

If p is prime and divides $n^b + a$, then either $p|a$, in which case $k = p$ is suitable for we have $p^b < n$ by (i), or else $p|((a^{v+mb})^s \pm 1)$, in which case we use (ii) and the lemma to conclude that

$$p^r < |a|^{vs+mbs} = \frac{n^b}{|a|} < n^b,$$

so that $p^b < n$ and a suitable k is given by $k \equiv n \bmod(p)$, $0 < k < p$. □

A Lower Bound

E 3025 [1983, 706]. *Proposed by S. W. Graham, Michigan Technological University, and D. Hensley, Texas A & M University.*

Let $\lambda(n)$ be the function on the positive integers determined by the conditions that $\lambda(p) = -1$ for all primes p and $\lambda(mn) = \lambda(m)\lambda(n)$ for all m and n. (This is commonly referred to as Liouville's lambda function.) Show that there exists a $C > 0$ such that the number of $n \leqslant x$ for which $\lambda(n) = \lambda(n+1)$ exceeds Cx, for x sufficiently large.

Solution by James Propp (graduate student), University of California, Berkeley. I claim that any $C < 1/4$ will do. Call n *good* if $\lambda(n) = \lambda(n+1)$ and *bad* otherwise. To prove the claim, note that when $\lambda(k) \neq \lambda(k+1)$ we also have $\lambda(2k) \neq \lambda(2k+2)$, so that either $\lambda(2k) = \lambda(2k+1)$ or $\lambda(2k+1) = \lambda(2k+2)$. Therefore each triplet $\{k, 2k, 2k+1\}$ contains a good n. Letting k range from 1 to $[(x-1)/2]$, we obtain at least $\frac{1}{2}[(x-1)/2]$ good $n \leqslant x$ (each n is contained in at most two triplets). Finally, we observe that $\frac{1}{2}[(x-1)/2] > Cx$ for all sufficiently large x.

Also solved by I. C. Bivens, R. Breusch, J. W. Grossman, S. H. Gould, J. Hook, L. R. King, O. P. Lossers (The Netherlands), J.-M. Monier (France), V. Pambuccian (Romania), University of South Alabama Problem Group, and the proposers.

Skew-Symmetric Matrix Decomposition

E 3026 [1983, 706]. *Proposed by H. Kestelman, University College, London.*

Let A be a skew-symmetric matrix of order $2n$ considered as a function of its $n(2n-1)$ independent upper off-diagonal elements; these are complex numbers ordered conventionally as $x_1, x_2, \ldots, x_N$. Show that there exist matrices F and G whose elements are rational functions of the x_r's with rational (real) coefficients, and which are such that $A = FG$ and $\det F = \det G$.

Solution by Thomas L. Markham, University of South Carolina. We use induction on n. For $n = 1$, A has order 2, and the result is easy to verify. Assume $A \neq 0$, and A has a proper leading principal submatrix A_1 of even order, which is invertible. (If no such A_1 exists, we consider PAP^T instead of A, where P is a permutation matrix. In particular, if $A \neq 0$, there is some upper-diagonal entry $a_{ij} \neq 0$. Applying a permutation to rows 1 and j, and columns 1 and j, we can take $A_1 = \begin{pmatrix} 0 & -a_{ij} \\ a_{ij} & 0 \end{pmatrix}$.) Partition A as $\begin{pmatrix} A_1 & A_2 \\ -A_2^T & A_3 \end{pmatrix}$. By induction, we write $A_1 = F_1G_1$ with the required properties. Note that $A_3 + A_2^TA_1^{-1}A_2$ is skew-symmetric of even order, and hence can be written as F_2G_2, where $\det F_2 = \det G_2$. Let

$$F = \begin{pmatrix} F_1 & 0 \\ -A_2^TG_1^{-1} & F_2 \end{pmatrix} \quad \text{and} \quad G = \begin{pmatrix} G_1 & F_1^{-1}A_2 \\ 0 & G_2 \end{pmatrix}.$$

It is clear that the elements of F and G are rational functions of the x_r's with rational coefficients, and $\det F = \det G$.

Also solved by J.-M. Monier (France) and the proposer.

A Geometric Factor

E 3030 [1983, 707]. *Proposed by R. E. Shafer, Berkeley, CA.*

Let $a_0 \geqslant a_1 \geqslant a_2 \geqslant \cdots \geqslant a_{N-1} > 0$. The polynomial $P_N(z) = \sum a_i z^i$ has zeros $|\zeta| \geqslant 1$. If for one of these zeros $|\zeta| = 1$, then there is $n \geqslant 2$, $n|N$, such that

$$P_N(z) = \frac{z^n - 1}{z - 1} \sum_{i=0}^{\frac{N}{n}-1} a_{ni} z^{ni}.$$

Solution by O. P. Lossers, Eindhoven University of Technology, The Netherlands. Since

$$(z-1)P_N(z) = a_{N-1}z^N + (a_{N-2} - a_{N-1})z^{N-1} + \cdots + (a_0 - a_1)z - a_0,$$

it follows from $P_N(\zeta) = 0$ and $|\zeta| = 1$ that

$$\begin{aligned} |a_0| &= |a_{N-1}\zeta^N + (a_{N-2} - a_{N-1})\zeta^{N-1} + \cdots + (a_0 - a_1)\zeta| \\ (*) \qquad &\leqslant |a_{N-1}\zeta^N| + |(a_{N-2} - a_{N-1})\zeta^{N-1}| + \cdots + |(a_0 - a_1)\zeta| \\ &= a_{N-1} + (a_{N-2} - a_{N-1}) + \cdots + (a_0 - a_1) = a_0, \end{aligned}$$

so the inequality $(*)$ is in fact an equality.

Since $a_0 > 0$, the complex numbers $a_{N-1}\zeta^N, (a_{N-2} - a_{N-1})\zeta^{N-1}, \ldots, (a_0 - a_1)\zeta$ must also be nonnegative. In particular $\zeta^N = 1$. Let n be the order of ζ. Then $n|N$ and $n \geqslant 2$ (for $P_N(1) \neq 0$). If $k \not\equiv 0 \pmod n$, then $\zeta^k \neq 1$ so $a_{k-1} - a_k = 0$. This shows that

$$P_N(z) = \frac{z^n - 1}{z - 1} \sum_{i=0}^{\frac{N}{n}-1} a_{ni} z^{ni}.$$

Also solved by A. Bondesen (Denmark), J.-M. Monier (France), and the proposer.

$\phi(n)$ Divides n Once Again

E 3037 [1984, 140]. *Proposed by Michael W. Ecker, Pennsylvania State University, Worthington Scranton Campus.*

Find all positive integers n such that $\phi(n)|n$.

Solution by Scott J. Beslin (student), University of Southwestern Louisiana. The case $n = 1$ is trivial and we need only consider the case $n > 1$, $n = p_1^{\alpha_1} \cdots p_t^{\alpha_t}$. If $\phi(n)|n$, then $n = x \cdot \phi(n)$ for some integer x. Thus

$$p_1 \cdots p_t = x(p_1 - 1) \cdots (p_t - 1).$$

Clearly, one of the primes p_j must be equal to 2 for the equality to hold. We can then write

$$2p_2 \cdots p_t = x(p_2 - 1) \cdots (p_t - 1).$$

It follows that n can contain *at most* a single odd prime, p_2. Let $p_2 - 1 = 2y$ for some integer y. Hence, $2p_2 = x(2y)$, which implies that $x = p_2$ and $y = 1$. Thus $p_2 - 1 = 2$ and $p_2 = 3$. Therefore $n = 2^b 3^c$, $b \geqslant 1$ and $c \geqslant 0$. Conversely, it is easily verified that if $n = 2^b 3^c$, then $\phi(n)|n$.

Also solved by 79 other solvers and the proposer.

W. C. Waterhouse notes that his solution to E 1575 [1964, 95] solves this problem also. R. E. Shafer notes that

his solution to 6070 [1977, 662] implies a solution to this problem. J. Suck located the problem in H. N. Shapiro's *Introduction to the Theory of Numbers* (New York, 1983) as Exercise 13 on page 76. Suck also pointed out that it appears in *An Introduction to the Theory of Numbers* by Ivan Niven and H. S. Zuckerman (4th edition, New York, 1980) as Problem 20 on page 51 with solution on page 322. Paul Peck and E. M. Klein also found the problem there, and David Boduch found it on page 255 of Niven and Zuckerman's third edition. W. I. Nissen found it in Oystein Ore's *Number Theory and its History* (New York, 1948) on page 115, without solution. Underwood Dudley points out that the answer to the problem is stated with a reference in W. Sierpinski's *Elementary Theory of Numbers* (Warsaw, 1964), page 232.

Norms and Means

E 3046 [1984, 369]. *Proposed by Steven C. Althoen and Lawrence D. Kugler, The University of Michigan, Flint, MI.*

It is easy to check that if $\|(x, y)\|_1$ and $\|(x, y)\|_2$ are norms on R^2, then so is their arithmetic mean $\frac{1}{2}(\|(x, y)\|_1 + \|(x, y)\|_2)$. Let

$$\|(x, y)\|_1 = (a^2x^2 + b^2y^2)^{1/2} \quad \text{and} \quad \|(x, y)\|_2 = (c^2x^2 + d^2y^2)^{1/2},$$

where a, b, c, and d are reals. Find necessary and sufficient conditions on a, b, c, and d so that the geometric mean $\sqrt{\|(x, y)\|_1\|(x, y)\|_2}$ will also be a norm.

Solution by the proposers. The answer is

$$\left|\frac{a^2d^2}{b^2c^2} - 17\right| \leqslant 12\sqrt{2}, \quad \text{or equivalently,} \quad 3 - \sqrt{8} \leqslant \left|\frac{ad}{bc}\right| \leqslant 3 + \sqrt{8}.$$

It is well known [1, p. 32] that $\| \ \|$ defines a norm in $\mathbb{R}^2$ if and only if the unit ball $\{x, y \mid \|x, y\| \leqslant 1\}$ is bounded and convex. We must find conditions that guarantee that

$$\sqrt[4]{(a^2x^2 + b^2y^2)(c^2x^2 + d^2y^2)} = 1 \quad \text{or} \quad (e^2x^2 + y^2)(x^2 + f^2y^2) = g^2$$

(where $e^2 = a^2/b^2$, $f^2 = d^2/c^2$ and $g^2 = 1/b^2c^2$) is convex. By symmetry it suffices to find necessary and sufficient conditions for which $y'' \leqslant 0$ for $x, y \geqslant 0$. Let $\alpha = e^2f^2 + 1$, then

$$-y' = \frac{x}{y} \cdot \frac{2e^2x^2 + \alpha y^2}{\alpha x^2 + 2f^2y^2},$$

$$-y'' = \left[2\alpha e^2u^2 + (12e^2f^2 - \alpha^2)uv + 2\alpha f^2v^2\right]K,$$

where $u = x^2$, $v = y^2$, $K = [(x^2 + y^2)N]/y^3[\alpha x^2 + 2f^2y^2]$, and $N = 2e^2x^2 + \alpha y^2$.

The discriminant of the quadratic form is $D = w^2[w^2 - 32w - 32]$, where $w = \alpha - 2$. Solving $D \leqslant 0$ yields

$$|e^2f^2 - 17| \leqslant 12\sqrt{2} \quad \text{or} \quad \left|\frac{a^2d^2}{b^2c^2} - 17\right| \leqslant 12\sqrt{2},$$

which is equivalent to

$$3 - \sqrt{8} \leqslant \left|\frac{ad}{bc}\right| \leqslant 3 + \sqrt{8}.$$

Reference

1. F. A. Valentine, Convex Sets, McGraw-Hill, 1964.

Also solved by W. A. Newcomb and R. S. Stevens.

Central Symmetry

E 3051 [1984, 438]. *Proposed by P. O'Hara and H. Sherwood, University of Central Florida.*

It is observed in plane analytic geometry that any set S that is symmetric with respect to both the x and y axes is also symmetric with respect to the origin. Does the statement remain valid if the y-axis is replaced by a line through the origin with inclination angle α?

Solution by David M. Wells, Pennsylvania State University, New Kensington. The statement remains valid if and only if $\alpha = (2k+1)\pi/2n$ for some integers k and n.

The symmetry conditions guarantee that a point with polar coordinates (r, θ) is in S if and only if S also contains the points $(r, 2n\alpha \pm \theta)$ for all integers n. These conditions force symmetry about the origin if and only if for any choice of θ, one of these points coincides with $(r, \theta + (2k+1)\pi)$ for some integer k. In particular, choosing $\theta = 0$, we must have $2n\alpha = (2k+1)\pi$. Conversely, if $\alpha = (2k+1)\pi/2n$, clearly $2n\alpha + \theta = \theta + (2k+1)\pi$.

Also solved by S. D. Bronn, D. Callan, Chico Problem Group, C. Chouteau, J. Dou (Spain), A. Facchini (Italy), J. Grossman, G. A. Heuer, R. T. Hood, M. Kantrowitz (student), L. Kuipers (Switzerland), N. J. Lord (England), D. Marcus, R. Mentock, E. Morgantini (Italy), W. A. Newcomb, J. Oman, J. Putz, P. J. Ryan, D. Spellman, G. Sylvester, U. of South Alabama Problem Group, Z. Usiskin, G. P. Wene, Western Michigan U. Problem Solving Group, and the proposers.

Several solvers mentioned that if S is closed, then the statement remains valid if α is an irrational multiple of π as well. There were twelve partial solutions which gave particular values for α and two incorrect solutions.

Optimal Strategies for a Computer Game

E 3060 [1984, 580]. *Proposed by Ira Gessel, Massachusetts Institute of Technology.*

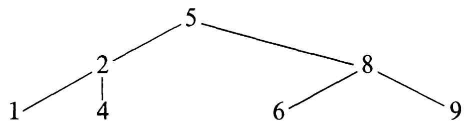
A popular computer game works as follows: The computer picks an integer from 1 to n at random. The player is given k opportunities to guess the number. After each guess the computer responds "correct," "too high," or "too low." For fixed k and n, find the probability of success with optimal play, and find all optimal strategies.

Solution by David Callan, University of Bridgeport, Connecticut. There is a perfect strategy if $n \leqslant 2^k - 1$. If $n > 2^k - 1$ an optimal strategy has a $(2^k - 1)/n$ probability of success and there are $\binom{n}{2^k - 1}$ such (deterministic) strategies—one for each selection of $2^k - 1$ integers from $\{1, 2, \ldots, n\}$.

Proof. With k guesses one has the opportunity to name (at most)

$$1 + 2 + 2^2 + \cdots + 2^{k-1} = 2^k - 1$$

numbers so one cannot improve on a $(2^k - 1)/n$ chance of success. For $n \geqslant 2^k - 1$ an optimal strategy is described by the tree graph below (case $n = 9$, $k = 3$)



Here the first guess is 5 and successive guesses (if necessary) proceed left or right according as the computer responds high or low.

If the random number occurs in such a graph the strategy will succeed provided that all numbers below a left (respectively, right) branch are less (respectively, greater) than the number directly above it. Exactly one such tree consisting of any given set of $2^k - 1$ numbers can be constructed: the median number is necessarily at the top, the $2^{k-1} - 1$ smallest numbers are below the left branch, so the median of these is in the $(2, 1)$ position, etc.

Thus there are $\binom{n}{2^k-1}$ optimal strategies when $n \geqslant 2^k - 1$ and a perfect strategy when $n = 2^k - 1$ (*a fortiori* when $n < 2^k - 1$).

Also solved by J. W. Grossman, C. Hurd, D. Knuth, O. P. Lossers (The Netherlands), W. A. Newcomb, and the proposer. Partially solved by L. R. King, M. Packter (South Africa), S. Singer, M. Vowe (Switzerland), J. T. Ward, and G. P. Wene.

ADVANCED PROBLEMS

For instructions about submitting solutions of these Advanced Problems, which should be mailed by February 28, 1987, see the inside front cover. The solver's full post-office address should be on each sheet.

6526. *Proposed by Grzegorz Rzadkowski, Agricultural University in Warsaw, Poland.*

Let f be a continuous real-valued function defined on the interval $[0, L]$. If there is an equilateral triangle ABC of sidelength L such that

$$f(AP) + f(BP) + f(CP) = 0$$

for all points P inside the triangle, prove that f is identically zero.

6527. *Proposed by Nicholas Strauss, Boston University.*

Let $J(m)$ be the $m \times m$ matrix whose (i, j)th entry is

$$\binom{i+j}{j}, \qquad 0 \leqslant i, j \leqslant m - 1.$$

Show that for all primes p and positive integers n, the matrix $J(p^n)$ is a cube root of unity modulo p.

6528. *Proposed by William P. Wardlaw, United States Naval Academy.*

The following is problem 11 on page 44 of C. Chevalley's *Fundamental Concepts of Algebra*, Academic Press (1956):

Let H be a subgroup of a group G. Show that the elements s of G such that the mapping $t \to sts^{-1}$ of G into itself map H into itself form a subgroup N of G, of which H is a normal subgroup; show that every subgroup of G containing H and in which H is normal is contained in N.

Provide an example to show that the assertion of the problem is false.

SOLUTIONS OF ADVANCED PROBLEMS

An Operator Fejér Kernel

6487 [1985, 63]. *Proposed by Pei Yuan Wu, National Chiao Tung University, Hsinchu, Taiwan, Republic of China.*

Let T be a contraction (i.e., $\|T\| \leq 1$) on a Hilbert space and, for $n = 1, 2, \ldots$, let

$$K_n = 1 + \sum_{k=1}^{n} \left(1 - \frac{k}{n+1}\right)(T^k + T^{*k})$$

be its "Fejér kernel." Prove by an elementary method (i.e., without the use of spectral theory or dilation theory) that $K_n \geq 0$ for $n = 1, 2, \ldots$.

Solution by Zachary Franco, student, Harvard University. Let X denote the Hilbert space. To

show that $K_n \geq 0$ is to show that $(K_n x, x) \geq 0$ for all x belonging to X. Since

$$(T^{*k}x, x) = (x, T^k x),$$

this is equivalent to $H_n(x) \geq 0$, where

$$H_n(x) = (n+1)(x,x) + \sum_{k=1}^{n} (n+1-k)\big[(T^k x, x) + (x, T^k x)\big].$$

Since $\|T\| \leq 1$, we have for $j = 0, 1, 2, \ldots$ that

$$(T^k x, x) \geq (T^{j+k}x, T^j x), \quad \text{and} \quad (x, T^k x) \geq (T^j x, T^{j+k}x).$$

Hence

$$H_n(x) \geq \sum_{j=0}^{n} (T^j x, T^j x) + \sum_{k=1}^{n} \sum_{j=0}^{n-k} \big[(T^{k+j}x, T^j x) + (T^j x, T^{k+j}x)\big]$$

$$= \left(\sum_{m=0}^{n} T^m x, \sum_{m=0}^{n} T^m x \right) \geq 0.$$

Two solutions, one by induction and one by means of an explicit matrix identity, were provided by Kooiti Hirano and Kyoko Kubo (Japan).

The proposer's solution used the operator identities

$$(n+1) K_n = S_n^* S_n + \sum_{j=0}^{n-1} (S_j^* S_j - T^* S_j^* S_j T)$$

and

$$((n+1) K_n x, x) = \|S_n x\|^2 + \sum_{j=0}^{n-1} \left(\|S_j x\|^2 - \|S_j T x\|^2 \right),$$

where

$$S_n = \sum_{k=0}^{n} T^k.$$

He also showed (by nonelementary methods) that if

$$p(z) = \sum_{k=-m}^{n} a_k z^k$$

is a trigonometric polynomial that is real and nonnegative on the unit circle, then $p(T) \geq 0$. Here

$$p(T) = \sum_{k=m}^{1} a_{-k} T^{*k} + \sum_{k=0}^{n} a_k T^k$$

by definition.

A Rayleigh Popular Problem

6488 [1985, 148]. *Proposed by Robert M. Young, Oberlin College.*

Show that if $\lambda_1, \lambda_2, \lambda_3, \ldots$ are the positive solutions of the equation $\tan x = x$, then

$$\sum_1^\infty 1/\lambda_n^2 = \frac{1}{10}.$$

Editor's Comment. Our readers located this problem in numerous sources. For example, Shafique Ahmed observes that Rayleigh's formula (from Watson's treatise on Bessel functions, 2nd ed., p. 502) asserts that

$$\sum 1/\lambda_{\nu,n}^2 = 1/4(\nu+1),$$

where the $\lambda_{\nu,n}$ are the positive zeros of the Bessel function $J_\nu(x)$ of order ν ($\nu > -1$). Since

$$J_{3/2}(x) = (2/\pi x)^{1/2}(x^{-1}\sin x - \cos x),$$

the result follows by taking $\nu = 3/2$. Had the editors been aware of this they would not have set forth the problem. However, several readers contributed insights into related problems that we feel are new, interesting, and at the very least not readily available in the literature. Moreover, the response by the readership was strong and varied (62 solutions received). This should justify the extended discussion below, that includes two solutions.

As for further references to the literature, readers mentioned treatises on heat conduction (H. S. Carslaw and J. C. Jaeger), on differential equations (G. F. Simmons), and M. R. Spiegel's paper "The summation of series involving roots of transcendental equations and related applications", Journal of Applied Physics (1953), 1103–1106. Many readers were inspired by the beautiful exposition of Euler's early discoveries in this area (he treated power series as if they were simply polynomials of very large degree), by G. Pólya in Chapter 2 of his book *Induction and Analogy in Mathematics* (Princeton, 1954). M. S. Klamkin and A. Meir note that the problem also occurs as the first part of Problem 77 of the Pi Mu Epsilon Journal (see the 1956 issue, pp. 187–188). The editor, however, believes that the best references for research on reciprocal power sums of Bessel zeros are Emman C. Obi, *Functional bounds, series, and log convexity of Rayleigh higher derivatives*, J. Math. Anal. Appl., 52 (1975), 648–659 and the references therein (especially to the work of N. Kishore).

Solution by R. William Gosper, Symbolics, Inc., Palo Alto, CA, and Eugene Salamin, Coherent, Inc., Palo Alto, CA (jointly). We prove a more general result. If $n \geqslant 2$ is an integer, if $p(z)$ is a polynomial, and if $\tan z = p(z)$ has nonzero roots $\{\lambda_j\}$ with corresponding multiplicities $\{r_j\}$, then

$$\sum_j r_j/\lambda_j^n = -a_{n-1},$$

where

$$\sum_k a_k z^k = \frac{(1-p'(z))\cos z + p(z)\sin z}{\sin z - p(z)\cos(z)} \overset{\text{def}}{=} g(z)$$

is a Laurent series.

Let $f(z) = \sin z - p(z)\cos z$. Then $f(z)$ has the same roots as $\tan z = p(z)$, and furthermore $f(z)$ has no poles. Let $n \geqslant 2$, $N > 0$ be integers, let R be a rectangle with vertices $\pm N\pi \pm iN$, let C be the boundary of R taken counterclockwise, and let

$$I = \frac{1}{2\pi i}\int_C \frac{f'(z)}{z^n f(z)}\,dz.$$

Here $|\tan z| < 2$ on C (details suppressed by the editor). Now, $f'(z)/f(z)$ equals $g(z)$, defined above, and after dividing numerator and denominator by $\cos z$, it is clear that for z on C, the integrand is $O(N^{-n})$ as $N \to \infty$. Since the length of C is $O(N)$, it follows that $I \to 0$ as $N \to \infty$ when $n \geqslant 2$.

By the residue theorem, we also have (possibly after deforming the contour to avoid poles)

$$I = \sum_{\lambda(j)\in R} \operatorname*{res}_{z=\lambda(j)} g(z)/z^n + \operatorname*{res}_{z=0} g(z)/z^n,$$

where the $\{\lambda(j)\} = \{\lambda_j\}$ are the nonzero roots of $f(z)$. If λ_j is a root of multiplicity r_j, then the residue at $z = \lambda_j$ is r_j/λ_j^n. The residue at $z = 0$ is the coefficient, a_{n-1}, of z^{n-1} in the Laurent series of $g(z)$. Taking the limit as $N \to \infty$ completes the proof.

In applying the above result, we may need to show that all roots of $h(z) = \tan z - p(z)$ have been found. A generalization of Rouché's theorem provides the needed help. Assume $p(z)$ is of degree $m > 0$ (since otherwise the roots of $h(z)$ are easily found). The poles of $h(z)$ are the same as the poles of $\tan z$, i.e., $(k + \frac{1}{2})\pi$ for some integer k. As $|z| \to \infty$, $|p(z)| \to \infty$. Hence if $h(\lambda) = 0$ and $|\lambda|$ is sufficiently large, λ must be near a pole of $\tan z$. Take the contour C as above, and let Z be the number of zeroes and let P be the number of poles, counting multiplicities, of $h(z)$ within C. Since C crosses the real axis at $k\pi$, a sufficiently large contour will avoid passing through a pole or zero. We have $Z - P = \Delta \arg(h(z))/2\pi$, the number of times $h(z)$ winds about the origin as z goes around C. However, for sufficiently large C, and z on C, we have $|\tan z| < 2 < |p(z)|$. Therefore, for large enough C,

$$Z - P = \Delta \arg(p(z))/2\pi = m,$$

the degree of $p(z)$. It also follows that for large $|k|$ there is exactly one root of $h(z)$ near $(k + \frac{1}{2})\pi$.

In the original problem, $p(z) = z$. Graphically it is clear that $\tan z - z$ has roots near its poles at $\pm 3\pi/2, \pm 5\pi/2, \ldots$. In addition, there are the two poles at $\pm\pi/2$ and the triple root at 0. By the above remarks, there are no more roots. Since the nonzero roots occur in pairs, $\lambda, -\lambda$, the sum of λ^{-2} over positive roots is half the sum over all nonzero roots. Since

$$g(z) = z \sin z/(\sin z - z \cos z) = 3z^{-1} - (1/5)z + \cdots,$$

the result follows.

This solves the stated problem, but the fun has just begun. The case $p = 2$ of another formula of Gosper-Salamin shows that for real $a \neq 1$ we have

$$\sum_{\lambda \neq 0} 1/\lambda^2 = \frac{3a - 1}{3(a - 1)},$$

where the sum is over all solutions of

$$\tan \lambda = a\lambda.$$

In fact, $\tan \lambda = \lambda$ has a triple root at 0 that spreads out on the real axis for $a > 1$ and on the imaginary axis for $a < 1$. The formula indicates that the cases $a = 1$ and $a = 1/3$ are the "most singular". (Before examining the latter, we add a remark of M. S. Klamkin and A. Meir. For $z = x + iy$ the equation

$$\sin z - az \cos z = 0$$

implies

$$e^{-2y}\left[(1 + ay)^2 + a^2x^2\right] = e^{2y}\left[(1 - ay)^2 + a^2x^2\right]$$

so the roots must again be real for $a < 0$).

For $a = 1/3$, the z term in the Laurent expansion of $g(z)$ vanishes, and the triple root at 0 has split into simple roots at 0, $i\mu$, and $-i\mu$, where $\mu > 0$ and

$$\tanh \mu = \mu/3.$$

Hence

$$\sum 1/\lambda_j^2 = 1/\mu^2 = (2.98470458535\ldots)^{-2},$$

where the sum is over $\lambda_j > 0$. Here μ has the continued fraction expansion

$$\mu = [b_0, b_1, b_2, \ldots] = [2,1,64,2,1,1,1,3,4,1,2,3,1,272213,1,2,1,16,\ldots];$$

the 13th partial quotient looks surprisingly large. The editor notes that the metric theory of continued fractions asserts that almost always

$$G_n = (b_1, \ldots, b_n)^{1/n} \to K = 2.685452001\ldots = \prod_{n \geq 1} \left(\frac{n+1}{n} \cdot \frac{n+1}{n+2} \right)^{\log_2 n},$$

the famous Khintchine constant. Here for $5 < n \leq 12$ we have $G_n < G_5 = 2.639\ldots$ while

$$G_{13} = 5.28\ldots \quad \text{and} \quad G_{14} = 4.69\ldots$$

are surprisingly large. For another point of view set

$$L_n(x) = \max\{a_j(x) : 1 \leq j \leq n\}, \quad 0 < x < 1.$$

Then with respect to the Gauss probability measure on [0, 1], namely

$$P(dx) = \frac{dx}{(1+x)\ln 2}$$

(see J. Galambos, *The distribution of the largest coefficient in continued fraction expansions*, Quart. J. Math. Oxford Ser. (2) 23 (1972), 147–151), we have

$$\lim_{N \to \infty} P\{L_N < yN/\ln 2\} = e^{-1/y}.$$

Again (e.g., set $N = 13$ and $y = 512$) the continued fraction for $\mu - 2$ seems most exceptional!

Solution by Klaus Zacharias, Berlin. We let sh(b) and ch(b) denote the hyperbolic sine and cosine of b, respectively. Now consider the (selfadjoint) Sturm-Liouville eigenvalue problem

$$-y'' + b^2 y = \mu y,$$
$$y(0) = 0, \quad y'(1) = y(1),$$

where $0 < x < 1$ and b is constant. For $\mu = 0$ obviously $y = 0$ is the only solution, so we can represent the solution (with the same boundary conditions) of

$$-y'' + b^2 y = f(x)$$

by

$$y(x) = \int_0^1 G(x,t) f(t)\, dt,$$

where the Green's function is

$$G(x,t) = \frac{\operatorname{sh}(bx)}{b}\left[\operatorname{ch}(bt) + \left(\frac{\operatorname{ch}(b) - b\operatorname{sh}(b)}{b\operatorname{ch}(b) - \operatorname{sh}(b)}\right)\operatorname{sh}(bt)\right]$$

for $0 \leq x \leq t \leq 1$ and

$$G(x,t) = G(t,x)$$

for $t \leq x$. This formula is hidden in T. Kato, *Perturbation Theory for Linear Operators*, Springer-Verlag, Berlin, 1980, Chapter 5, Example 4.14, where more information on the self-adjointness and spectrum of our differential equation may be found. By classical Hilbert-Schmidt theory

$$G(x,t) = \sum_{n=0}^{\infty} \frac{v_n(x) v_n(t)}{\mu_n},$$

where the μ_n are the (simple) eigenvalues with corresponding normed eigenfunctions v_n. By the trace formula,

$$\int_0^1 G(x,x)\, dx = \sum_{n=0}^{\infty} 1/\mu_n = b^{-2} + \sum_{n=1}^{\infty} 1/\mu_n.$$

Upon setting

$$y = A\cos(\mu - b^2)^{1/2}x + B\sin(\mu - b^2)^{1/2}x,$$

we find $A = 0$ and $\mu = \lambda^2 + b^2$, where λ is positive and

$$\tan\lambda = \lambda.$$

Thus our formula for $G(x, t)$ yields

$$b^{-2} + \sum_{j=1}^{\infty} \frac{1}{\lambda_j^2 + b^2} = \frac{1}{4b^2}[(\mathrm{ch}(2b) - 1) - (\mathrm{ch}(b) - b\,\mathrm{sh}(b))]\left[\frac{2b - \mathrm{sh}(2b)}{b\,\mathrm{ch}(b) - \mathrm{sh}(b)}\right].$$

By elementary asymptotics this reduces to the desired identity when $b \to 0$.

Approaches similar to Zacharias' led many to

$$\sum_{n=1}^{\infty} \frac{\sin\lambda_n x}{\lambda_n^2 \sin\lambda_n} = \frac{5x^3 - 3x}{20}$$

(consider $x = 1$), and S. K. Rangarajan also to

$$\sum_{n=0}^{\infty} \frac{\sin\phi_n x}{\sin\phi_n} \frac{1}{\phi_n^2 + b^2 + b} = \frac{x}{2(b+1)},$$

where $b \neq -1$ and the ϕ_n are the positive roots of

$$\phi\cos\phi + b = 0.$$

For a detailed study of the Bessel-zeta function

$$z(\nu, s) = \sum_{n=1}^{\infty} 1/\lambda_{\nu,n}^s,$$

where s and ν are complex variables, some relevant eigenvalue problems, and further references see John Hawkins, Doctoral Thesis, University of Illinois (Urbana), 1983 (unpublished). A partial summary of his work is given in K. B. Stolarsky, *Singularities of Bessel-zeta functions and Hawkins' polynomials*, Mathematika, 32(1985), 96–103. Finally, P. Barrucand (France) writes that many interesting formulas connected with this problem are in an old and forgotten paper by A.R. Forsyth, *The expression of Bessel functions*..., Messenger Math., 50(1921), 129–149.

The 62 solutions received may be loosely classified as follows: contour integration and (Hadamard) infinite product expansions (45), Sturm-Liouville techniques (12) and manipulation of Bessel functions (5).

---

175. MISCELLANEA

DO AS I SAY, NOT AS I DO
OR
HOW TO STATE A RULE IN A STATEMENT THAT BREAKS THE RULE

(From the entry "or" in the American Heritage Dictionary of the English Language)

When the elements [connected by *or*] do not agree in number, or when one or more of the elements *is a personal pronoun*, the verb is governed by the element to which it is nearer.

—J. CHRIS FISHER, University of Regina

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER

Basic Algebra I, second edition. By Nathan Jacobson. W. H. Freeman, 1985. xvii + 492.

ANDY R. MAGID
Department of Mathematics, University of Oklahoma, Norman, OK 73069

What is Algebra and how much of it does the mathematical professional need to know? The Algebra of which we speak is the "Modern," or "Higher," or "Abstract" Algebra of the upper division and post graduate curriculum (although we need to be careful when speaking with the non-mathematical public: a colleague from the social sciences once responded to my announcement that my research was in Abstract Algebra by inquiring "Isn't that redundant?"), and it consists of the study of groups, rings, and fields.*

Such studies are a necessary part of the training of prospective group theorists, and ring theorists, and algebraic geometers, and so forth, but why should control theorists or combinatorialists also be expected to learn (as the book under review titles it) *Basic Algebra*? A first answer is the utilitarian one: because they also need to use Algebra. Tannenbaum's applications of invariant theory in control theory [**13**] and Stanley's applications of ring theory in combinatorics [**12**] are typical examples. But like most curricular issues, the answer of relevance alone is not entirely satisfactory. One feels there is something else, some algebraic way of thinking which transcends the contents of the Basic Algebra course, and which the prospective mathematician needs as much as or more than he needs the applicable facts about algebraic structures that he will see defined and proved.

This reviewer will not attempt a definition of this essence of algebraic thought. But perhaps the reader will excuse a little soft speculation. Algebra seems to be about the holistic properties of collections of things which, while they have no independent status, derive their significance from the relations and operations that exist on the collection as a whole. It makes little difference that the elements of our ring, say, are matrices, or differential operators, or formal linear combinations of group elements; in fact it can even be a positive hindrance to think of them that way. For example, consider the groups of Galois Theory: it is vital, in the end, to think of these groups as groups of substitutions in the roots of the equation being examined, but the technical problems which would attend an attempt to prove the criterion for solvability by radicals while working exclusively within this particular representation would be enormous. Far better to ignore the nature of the elements and to think of the group as a thing in itself. To take another example: the notion that the cosets of a normal subgroup of a group, while they have intrinsic meaning as subsets of the original group, are best thought of as unities, as elements of a new group, the quotient group, is often the *pons asinorum* of the Basic Algebra course. Those who cross it successfully usually do learn to think algebraically. It is probably unfair to claim this thought mode—ignoring the essence of elements of a structure and focusing on their relations—as exclusive to Algebra. This is the basis of much modern abstraction, and not only in mathematics; see for example [**11**]. But Algebra does seem to appear whenever structures dominate a piece of mathematical thought.

From soft speculation we turn to the hard questions of curriculum. Granted that Basic Algebra is important and that all mathematicians need to know it, what exactly is to be learned and how do we teach it? We teach, of course, as we were taught, which for the American Algebra

*I am reminded also of another non-mathematical friend, who learned of Algebra while helping a relative perfect the grammar in some papers on commutative rings, and who related to me some metaphysical concerns about people who were always "Letting something be something else," for example "Let R be a ring ... ."

professoriate of today means primarily from the undergraduate texts by Birkhoff and Mac Lane [2] and Herstein [5], the graduate texts by Van der Waerden [14] and Lang [10], and the first edition of *Basic Algebra* and its predecessors [7], [8], [9]. Thus we begin our courses with group theory (especially finite groups), continue with ring theory (especially polynomial rings in a single variable over a field), and then end with field theory (especially the Galois theory of fields generated by roots of a polynomial). It's hard to find fault with this tradition: the results of the early lectures, such as Lagrange's theorem about the order of a subgroup of a finite group dividing the order of the group, play a significant role in the last ones, when the correspondence between subfields and subgroups is established in the Galois theory. And the path has been so well smoothed by our pedagogical predecessors that there's even a month or so at the end of the year to lecture on additional topics, such as representations of finite groups or rings of algebraic integers.

Generations of students have thus come to see the Galois theory as the paradigm of Algebra: the hard problem (solvability of the quintic or higher degree equation) converted to an easier one (solvability of the Galois group). Not a bad approach to a problem, if it can be pulled off. But is it truly *the* paradigm? Suppose instead we had taught Stone's theorem, which realizes Boolean algebras as algebras of $\{0,1\}$-valued functions, and points the way to the notions of spectrum in such areas as the maximal ideal space in Banach algebras, algebraic geometry, and elsewhere. Or suppose we had lectured on the Khron-Rhodes theorem realizing finite semigroups in wreath products of finite simple groups and flip-flops, and which would provide an entry point to the theory of automata and formal languages (see [4]). It might be possible to use Bass's lectures on K-theory and linear groups [1] as a skeleton on which to hang a Basic Algebra course, thereby introducing students to combinatorial group theory, the commutative algebra of Noetherian rings, and functors. There have been some efforts to produce Algebra books and courses along such non-traditional lines (although not, I think, along these three), but they are primarily intended for non-mainstream Algebra. Examples of this genre are Childs' interesting classical algebra book [3], and the various "applied algebra" texts.*

For traditional Algebra, Jacobson's *Basic Algebra I* is hard to beat. It's clear and complete, and it studs the path of groups through rings to fields and beyond with all sorts of mathematical jewels, such as the characterization of symmetric polynomials, or the identification of the composition algebras. Topics covered include Sylow Theorems, free groups, Euclidean Domains, quaternions, finitely generated modules over PID's, ruler and compass constructions, Galois theory, trancendance of e and π, real fields, classical groups, simplicity of PSL_n, real division algebras, and the fundamental theorem of projective geometry. There's no doubt that this is basic Algebra, and all mathematicians need to know it. Moreover, they can learn it from *Basic Algebra I*: there is attention here to exposition and detail that students (and their teachers) will appreciate.

Mathematical developments in Algebra since the first edition of 1974—such as the classification of the finite simple groups, the freedom of projective modules over polynomial rings, or the finite generation of groups of integral points on Abelian varieties—are not in this text (properly they would be much closer to the topics of the companion volume *Basic Algebra II*), but some changes have been made, improving the chapters on Galois theory and Elimination theory. The text was completely reset for this edition, with the attendant risks (although I only found one new typo: the inequality sign in the statement of Theorem 3.9 of Chapter 3 has been incorrectly printed as an equality for this edition). The publisher's blurb which accompanied my review copy says that "a solutions manual accompanies the text". I haven't seen it, and in fact my only

*Another worthy experiment that deserves some notice is Hochschild's advanced remedial Algebra book for graduate students [6], which makes a serious effort in its exercise sets to use the mathematics of computation and algorithmic thinking.

evidence for its existence is this single inscription, but this seems like such a wonderful idea that the publisher and author and whoever else may be responsible for this innovation deserve high praise. Would that every text at this level would be so accompanied. (Perhaps also we should comment on the level of this text: the author lists only linear algebra as prerequisite; when we use this book as a text at my university, it is for our first year graduate course which has a year of undergraduate algebra as a prerequisite.)

So what is Algebra and how much should one know? *Basic Algebra I* is an excellent answer.

References

1. H. Bass, Introduction to Some Methods of Algebraic K-theory, CBMS Reg. Conf. Series in Math. #20, Amer. Math. Soc., Providence, 1974.

2. G. Birkhoff and S. Mac Lane, A Survey of Modern Algebra (revised edition), Macmillan, New York, 1953.

3. L. Childs, A Concrete Introduction to Higher Algebra, Springer-Verlag, New York, 1979.

4. S. Eilenberg, Automata, Languages, and Machines, vol. B, Academic Press, New York, 1976.

5. I. Herstein, Topics in Algebra, Blaisdell, Waltham, MA, 1964.

6. G. Hochschild, Perspectives of Elementary Mathematics, Springer-Verlag, New York, 1983.

7. N. Jacobson, Lectures in Abstract Algebra, vol. I, Basic Concepts, Van Nostrand, New York, 1951.

8. ______, Lectures in Abstract Algebra, vol. II, Linear Algebra, Van Nostrand, New York, 1953.

9. ______, Lectures in Abstract Algebra, vol. III, Theory of Fields and Galois Theory, Van Nostrand, New York, 1964.

10. S. Lang, Algebra, Addison-Wesley, Reading, MA, 1965.

11. J. Piaget, Structuralism, Basic Books, New York, 1970.

12. R. Stanley, Combinatorics and Commutative Algebra, Birkhauser, Boston, 1983.

13. A. Tannenbaum, Invariance and Systems Theory, Lecture Notes in Math #845, Springer-Verlag, New York, 1981.

14. B. van der Waerden, Modern Algebra, Fredrick Ungar, New York, 1949.

Exercises in Number Theory. By D. P. Parent. Springer-Verlag, New York, 1984, x + 532 pp.

BRUCE C. BERNDT

Department of Mathematics, University of Illinois, Urbana, IL 61801

On the first page of Emil Grosswald's *The Theory of Numbers*, the author remarks, "In number theory, on the other hand, one can ask many questions in such simple terms, that the famous 'man in the street' can immediately understand but generally not answer them!" The book under review is not designed for the proverbial "man on the street". But nonetheless, the quote above and *Exercises in Number Theory* point to the primary source of mathematics—questions and problems. Many a problem "withereth as the grass;" for example, a problem may be trivial, unimportant, meaningless, or inelegant, and so is soon forgotten. Others like Goldbach's conjecture and the Riemann hypothesis remain open for decades, possibly even centuries.

Those of us who have chosen careers in mathematics were probably turned in this direction by interesting and challenging problems. Stimulating problems give vent to our basic desire to be creative, ingenious, and clever. At times, this basic need has been thwarted in the classroom at the expense of a truckload of definitions, theorems, and proofs with few if any stimulating problems suggested to the students. Fortunately, in the past few years, this sterile approach has been losing its footing. Several "problems" books have entered the market. However, it is highly likely that the most influential book of this genre will continue to be Pólya and Szegő's masterful two volumes, *Problems and Theorems in Analysis*. Springer Verlag has sold over 8000 copies in the German edition and 4000 in the English translation. By the *New York Times* best seller's list, such

figures are not very impressive, but these numbers far outdistance sales for comparable books at a similar level.

The present book by D. P. Parent is considerably more specialized than most books of this sort. (D. P. Parent is a pseudonym for twelve French authors.) The problems are best suited for advanced graduate courses in number theory. Most teachers interested in stimulating students with challenging problems in number theory would do better to consult such books as Niven and Zuckerman's *An Introduction to the Theory of Numbers* or Sierpinski's *250 Problems in Elementary Number Theory*.

Parent's book consists of about 150 exercises, many developing an idea or theory and containing several parts, and unevenly divided among ten topics including arithmetic functions, rational series, distribution modulo 1, transcendance, p-adic analysis, and modular forms. Very little preliminary exposition is given. Thus, this book best serves as a supplement to other reading or to an advanced course. References to several books are given, but no research papers are cited although several problems are lifted from original papers. Complete, detailed solutions to all problems are provided.

This book is not meant for casual perusal in the hope of picking up a gem here and there. However, the serious practitioner of number theory will find much here that is enriching.

A First Course in Differential Geometry. By Izu Vaisman. Marcel Dekker, New York, 1984. v + 169 pp.

David A. Singer
Department of Mathematics, Case Western Reserve University, Cleveland, OH 44118

The presence of this new book inspires two questions: What is or should be contained in a first course in differential geometry, and do we really need another textbook about it? The recent explosion of interest in the subject and its connections with partial differential equations and with geometric topology, together with the increased use of geometric techniques in engineering (e.g., systems engineering) and physical sciences, has created a (welcome) demand for courses in geometry; this suggests an affirmative answer to the second question. If the course is to be aimed at an audience including undergraduates and perhaps students who are not mathematics majors, then the topics must be chosen with some care.

The unifying idea which gives overall structure to a first course in differential geometry is surely that of curvature. For the student whose experience with geometry is limited to the Euclidean kind, this is the concept which captures the imagination. Indeed, if we consider the influence of the development of cartography, sparked by the need to explore the New World, on geometric thinking, we see that capturing the essence of curvature in a mathematical way was the driving force behind much of the work of Gauss and Riemann, the giants of the field.

Before Gauss, Leonhard Euler had dealt with the idea of the curvature of a surface through examination of plane sections. To describe the way a surface curves at a point P, one can form the intersection of it with a plane containing P and the normal vector $N(P)$ and then compute the curvature k of the intersection curve, which is of course planar. Euler showed [E] that each of these normal sections has curvature given by the formula

$$k = k_1\cos^2\theta + k_2\sin^2\theta,$$

where k_1 and k_2 are the maximum and minimum curvatures and θ is the angle between the plane corresponding to k_1 and that for k. The planes corresponding to k_1 and k_2 are orthogonal. Thus

Euler established that in some sense the curvature of a surface (at a point) could be described by means of two numbers. (Actually, this statement is more precisely due to Meusnier, who showed that even the curvatures of *non-normal* sections are determined by this pair of numbers.)

The modern theory of surfaces originates in the magnificent work of C. F. Gauss, who set forth many of the central ideas in his 1827 Gottingen lecture, "Disquisitiones Generales Circa Superficies Curvas" [**G**].

Gauss began by describing the metric structure of a surface in terms of what we now call the *metric tensor*. Endowing a portion of the surface with coordinates p, q, he described the measurement of length and angle in terms of three functions E, F, and G of p and q. These quantities are the "local coordinates" representation of the first fundamental form of the surface.

If M is an oriented surface in R^3, then the correspondence which assigns to each point P of M the unit normal vector $N(P)$—translated to the origin to give a point in the unit sphere S^2—is now known as the *Gauss map*. Using this function, Gauss defined the *integral curvature* of a region of M to be the (oriented) area of its image in S^2 under the Gauss map. Then he defined the *measure of curvature* $K(P)$ at a point P as the limiting value of the ratios of integral curvature to area for neighborhoods of P. This he related to the curvatures k_1 and k_2 by showing that the curvature $K(P)$ is equal to their product. This gives a very pretty geometric interpretation of curvature; in particular one can see the qualitative difference between positive and negative values of K manifested in the shape of the surface.

At this point Gauss proved a very elegant and remarkable result, one which he himself referred to as "a most excellent theorem", or *Theorema Egregium*. He derived a formula expressing $K(P)$ in terms of the functions E, F and G and their first two derivatives. Now if the surface is bent in such a way that no stretching or compressing takes place, then the quantities E, F and G do not change. Therefore, Gauss' Theorem tells us that curvature is a bending invariant.

What is so important about this result? Well, for example, since the curvature of the plane is identically 0, while the surface of the sphere of radius A is $1/A^2$, it follows immediately that no portion of a sphere can be flattened onto a plane without distortion. While this was well known to map-makers even in Ptolemy's time, Gauss provided the rigorous proof (and without the assumption of constant curvature). Later, Riemann would generalize this result to higher dimensions, showing that the geometry of Euclid is completely characterized (locally) by the vanishing of the (Riemann) curvature.

Of greater significance, Gauss introduced in this theorem the concept of "intrinsic" geometry of surfaces. One defines an intrinsic property of a surface to be one which depends only on its metric structure and not on the way it sits inside Euclidean space. Thus the principal curvatures at a point are *extrinsic* as is their sum (the *mean curvature*), while (by the *Theorema Egregium*) their product is intrinsic.

To understand the concept of intrinsic properties of surfaces, it is helpful first to investigate the properties of curves. The key idea here is that a regular curve (that is, a curve in R^3 with non-vanishing velocity) can be reparametrized to have unit speed. In terms of this canonical arc-length parameter s, the curvature function $k(s)$ and the torsion function $\tau(s)$ turn out to be a complete set of geometric information, in that the following theorem holds: If two curves have the same curvature and torsion at points with corresponding arc-length parameter, then one can be made to coincide with the other by a rigid motion of space. This follows from the Picard Theorem of ordinary differential equations applied to the *Frenet formulas*; these are the differential equations relating the coordinates of the curves. The curvature and torsion appear as coefficients of the equations. It also follows from the Picard theorem that any differentiable function $k(s) > 0$ and continuous function $\tau(s)$ arise as the curvature and torsion of a space curve.

In the preceding discussion the arc-length parameter was the intrinsic information, while the curvature and torsion were extrinsic. For surfaces the situation is much more interesting. As we

have mentioned, Gauss found a formula, now known as the *Gauss equation*, relating curvature to the metric tensor. There is another set of equations relating intrinsic and extrinsic information about the surface, known as the *Peterson-Mainardi-Codazzi* equations. These relations, discovered independently by K. M. Peterson in 1853, G. Mainardi in 1856, and D. Codazzi in 1867, combine with the Gauss equation to yield the *integrability conditions* of a surface. To understand this statement, let us look at the *second fundamental form* of a surface.

Recall that the first fundamental form of a surface measures infinitesimal lengths and angles. In other words, it is an inner product structure on each tangent space to M, varying "smoothly" with the point p in M. The functions E, F, and G are the entries in a positive-definite symmetric matrix which represents the inner product in the local coordinates of the surface. The second fundamental form is also a symmetric bilinear function on vectors in the tangent space, represented by a symmetric matrix with entries L, M, and N which are functions of the local coordinates. The simplest way to define it is to assume that the surface is tangent to a plane at the point p, so that the surface locally is the graph of a function whose Taylor series begins

$$f(u,v) = \frac{1}{2}(Lu^2 + 2Muv + Nv^2) + \cdots .$$

The quantities L, M, and N are extrinsic, but the ratio

$$(LN - M^2)/(EG - F^2)$$

is equal to the Gaussian curvature, an intrinsic quantity.

The PMC equations are formulas relating the quantities L, M, N, E, F, G and their derivatives to each other. These equations, together with the Gauss equation, form the necessary and sufficient conditions for those six functions to be the coefficients of the first and second fundamental forms of a surface in R^3. Given a vector valued function

$$X(u,v) = (x(u,v), y(u,v), z(u,v))$$

defining a piece of a surface, the functions $E(u,v)$, $F(u,v)$, and so on can be expressed in terms of the partial derivatives of X. The problem of finding a surface with specified data E, F, etc., is that of solving a first-order system of partial differential equations. Now instead of the Picard theorem we need the Frobenius theorem, which assures the existence and uniqueness of a solution provided certain integrability conditions are satisfied. These conditions are precisely the Gauss and PMC equations.

The functions E, F, G, L, M, N contain all of the (local) geometric information about the surface as it sits inside Euclidean space. The first three functions contain all of the intrinsic geometric information about the surface. The Gauss and PMC equations tell which extrinsic information about a surface is forced on it by its intrinsic geometry.

The purpose of Izu Vaisman's well-written book is to present the ideas I have described above to upper-level undergraduates in a one-semester course, and in addition to give the students a hint of the kind of problems that differential geometers look at. There are, of course, many excellent books already available to instructors, of which I will just mention a couple of my favorites.

Foremost is the Great American Differential Geometry Book of Michael Spivak [**Sp**]; I advise my students to buy all five volumes and read them, but in one semester one must either skip volume one (!) or never get to the beautiful volume two. (Someday I would love to teach a two-year course from that book.) Swinging to the other extreme, there is that magnificent marvel of conciseness, *Notes on Differential Geometry* by the late Noel Hicks [**Hi**]. This book (unfortunately out of print) is marvelous for graduate students who want to take a fast jet to the frontier of research, but is too terse for most undergraduates. Other favorites of mine include: the classic by Struik [**St2**], with its old-fashioned notation and many goodies not found in more modern

books; doCarmo's book [**dC**], which has a modern approach and has been my most recent choice as a textbook; and Stoker's book [**St1**], which is extremely readable and particularly appealing in classes with Physics majors. I should also mention the classic lecture notes of Hopf [**H**].

Vaisman's book is less extensive than the books mentioned above, since the author's stated goal is to give an introductory course for advanced undergraduates. He begins by carefully developing the notion of abstract manifolds and tensors. This makes good sense in a class with students in physics or engineering who have already had to deal with local coordinates. My personal preference when teaching mathematics majors is to begin with examples of curves and surfaces and delay the abstract formulation, as is commonly done (for example in [**M**], which delays the notion of manifold until the final chapter). Vaisman chooses to formulate things in terms of tensor notation ("the debauch of indices"), but he also uses vector notation, frequently shifting from one to the other.

The second part of the book treats the subject of smooth curves in the plane and in R^3. Here he develops the basic concept of the moving frame along a curve, the Frenet equations, and the fundamental existence and uniqueness theorem. He also includes some interesting and unusual additional material: he classifies the local geometric behavior of curves with isolated singular points of finite "order", "class" and "rank". He also proves Fenchel's theorem that the total absolute curvature of a closed curve is at least 2π, with equality only for convex planar curves.

In the main part of the book, Vaisman develops the basic theory of surfaces much as described above, including parallel translation and the equations of a geodesic.

The remainder of the book discusses special topics. The most significant one from the historical perspective is the local classification of surfaces of constant curvature. Gauss had shown that in order for two surfaces to be *isometric* (have the same metric structure) they must have the same Gauss curvature everywhere. In the case of constant curvature, the converse holds, as was proved by Minding: if the curvature of a surface vanishes, it is locally isometric to the plane. (Riemann later generalized this result to arbitrary dimensions, using Riemannian curvature.) When K is constant and positive it is locally isometric to a sphere; when K is constant and negative it is locally isometric to a "pseudosphere". This connects differential geometry with the classical geometries of Euclid, Riemann, and Lobachevskii and reveals it to be a common generalization of them. Curvature is the unifying geometrical concept.

A final bonus is a proof of the Gauss-Bonnet theorem, the famous result that says that the total Gauss curvature of a compact surface is a *topological* invariant. It is a most appropriate way to end a first course in differential geometry, with a theorem bearing the name of its illustrious parent.

There are no doubt many other great ideas in differential geometry that one might like students to see in a first course, e.g., rigidity theorems or perhaps the Bonnet-Myers theorem. Vaisman has quite sensibly limited the content to what one can reasonably expect to do in a one-semester course. His book should prove a useful text for such a course.

References

[dC]. M. P. Do Carmo, Differential Geometry of Curves and Surfaces, Prentice-Hall, Englewood Cliffs, 1976.

[E]. Leonhard Euler, Recherches sur la courbure des surfaces, Histoire de l'Académie Royale des Sciences de Berlin, 16 (1760) 119–143.

[G]. Carl F. Gauss, General Investigations of Curved Surfaces, Raven Press, New York, 1965. (English translation.)

[Hi]. Noel Hicks, Notes on Differential Geometry, Van Nostrand Reinhold, London, 1971.

[H]. Heinz Hopf, Lectures on Differential Geometry in the Large, lecture notes, Stanford University, 1955.

[M]. R. S. Millman and G. D. Parker, Elements of Differential Geometry, Prentice-Hall, Englewood Cliffs, 1977.

[Sp]. Michael Spivak, A Comprehensive Introduction to Differential Geometry, Publish or Perish, Boston, 1975.
[St1]. J. J. Stoker, Differential Geometry, Wiley-Interscience, New York, 1969.
[St2]. Dirk Struik, Lectures on Classical Differential Geometry, Addison-Wesley, Reading, 1950.

LETTERS TO THE EDITOR

For instructions about submitting letters for publication in this department see the inside front cover.

Editor:

In beginning his review of Martin Davis' *Computability and Unsolvability*, Stob [**1**] states the following: "... To see the issues involved, consider the knapsack problem. (Theoretical computer scientists love cute names for problems. The Dining Philosophers and the Byzantine Generals are two recent examples.)" Most readers of the MONTHLY would not have any difficulty with these statements and the implication that the knapsack problem is from the domain of computer science and was so named by some (unknown) computer scientist. I read them and was bothered, hence these comments.

It seems as if computer scientists are discovering (and rediscovering) that the field of operations research, and the related areas of applied mathematics, optimization and combinatorics, have originated most of the problems that form the basis of that field's treatment of computational complexity. Many of the interesting problems in complexity originated from linear programming and the simplex algorithm, the traveling salesman problem, among others, and of course, the knapsack problem; see, for example, *Combinatorial Optimization: Algorithms and Complexity* by Papadimitriou and Steiglitz, Prentice-Hall, 1982.

The knapsack problem was first named and discussed by G. B. Dantzig (the originator of linear programming) in his RAND Corporation Report "Notes on Linear Programming: Part XXXV—Discrete-Variable Extremum Problems," December 6, 1956, and printed in his book *Linear Programming and Extensions*, Princeton University Press, 1963. There, Dantzig defines the knapsack problem in a most pedagogically satisfying way (anyone can understand the problem statement), describes various algorithms for solving it, and notes that it can also be treated by Bellman's dynamic-programming functional-equation procedure.

The purpose of my letter is to stress to those readers of the MONTHLY, who are interested in learning and teaching about computational complexity and its origins, that they should first delve into the extensive, interesting, fascinating, mathematically sophisticated world of operations research.

Reference

1. M. Stob, Book Review: Computability and unsolvability, this MONTHLY, 93 (1986) 69–71.

Saul I. Gass
College of Business and Management
University of Maryland
College Park, MD 20742

If you teach math, can you pass this simple true or false test?

100%

1 John Kemeny and Tom Kurtz know how to teach mathematics.

[redacted] ☐ **False.**

John Kemeny headed the math department at Dartmouth College. He's also President Emeritus of Dartmouth, and internationally known as a speaker, teacher, and philosopher. Tom Kurtz helped shape the world of academic computing. Now you can let them help teach your class algebra, trigonometry, pre-calculus, calculus, probability theory, statistics, and discrete mathematics.

2 Software in the Kemeny and Kurtz Mathematics Series will fit into any teacher's individual classroom program.

[redacted] ☐ **False.**

This software isn't tied to any particular text or course. So it can easily supplement your own curriculum at your own speed.

3 Computer training in mathematics classrooms is now affordable.

[redacted] ☐ **False.**

They're reasonably priced at only $49.95 per single copy. Generous quantity discounts are available, allowing every student access to the learning power of these products.

4 *Classroom Computer Learning* named the True BASIC Language System™ one of the "Outstanding Products of 1985."

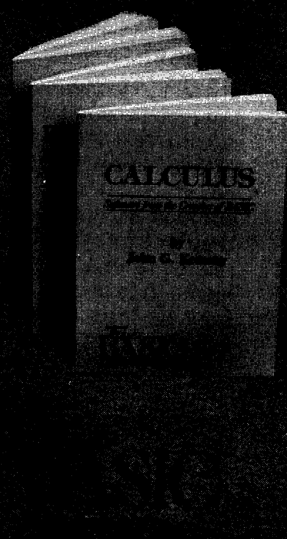
[redacted] ☐ **False.**

Not only that, but *Byte* magazine called it "Superior to Microsoft™ BASIC." *PC* magazine said it was "the easiest to learn of all the BASICs. . . ." And finally, *Electronic Learning* concluded "Good graphics have never been easier . . . only $149.90." And the same thoughtful design that went into True BASIC has gone into the Kemeny and Kurtz Mathematics Series, too.

O.K. You've made me a True believer.

You should consider the advantages of the Kemeny and Kurtz Mathematics Series for your classes. It's a flexible, affordable tool to help your students explore the concepts of mathematics. True BASIC. You can depend on us, 100 percent. The Kemeny and Kurtz Math Series is available for the IBM™ PC and compatibles, Apple Macintosh,™ and Commodore Amiga.™

To order, call 1-800-TR-BASIC.
In New Hampshire,
call **(603) 643-3882.**
Or send the coupon below.



BASIC information necessary to pass this test: John Kemeny and Tom Kurtz, co-inventors of the original BASIC programming language, have now developed a series of software programs designed to deepen students' understanding of the principles of mathematics.

System requirements: IBM PC 192K and color graphics adaptor, DOS 2.0+; Macintosh 512K; Amiga 512K

30 Day, No-Risk Trial

Act now to take advantage of this special offer. Order today, use the programs for 30 days, and if you're not completely satisfied, send them back for a full refund. My computer is: ☐ IBM PC/compatible ☐ Apple Macintosh ☐ Commodore Amiga

☐ I want to order the following Kemeny and Kurtz Mathematics programs: ☐ Algebra ☐ Trigonometry ☐ Pre-Calculus ☐ Calculus ☐ Probability Theory ☐ Statistics ☐ Discrete Mathematics ☐ The True BASIC Language

☐ Send me True BASIC product information.

Name________________ Title________________

Company/University________________

Address________________

City, State, Zip________________

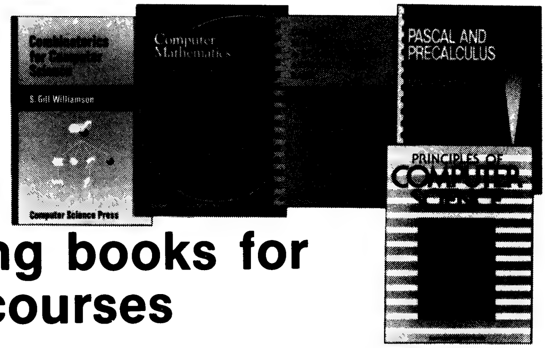
Telephone________________

Credit Card, Number,________________

Expiration Date________________

AMM-10

COMPUTER SCIENCE PRESS



has outstanding books for mathematics courses

COMBINATORICS FOR COMPUTER SCIENCE, ***S. Gill Williamson***

The two major subdivisions of combinatoric mathematics for computer science, enumeration and graph theory, are the subject of this outstanding new book, with emphasis on the conceptual needs of computer science. The book is an excellent choice for graduate students and advanced undergraduates in mathematics, applied mathematics, and computer science. ***1985, 479 pages, 0-88175-020-4, $39.95.***

INTRODUCTION TO COMPUTER MATHEMATICS, ***Russell Merris***

The relationship between computers and mathematics is the focus of this outstanding book which uses the computer as a tool in the study of probability, statistics, algebra, and geometry. The BASIC programming language is introduced as needed. For undergraduate courses. ***1985, 284 pages, 0-88175-083-2, $27.95; Teacher's Edition, 0-88175-112-X, $32.95; Teacher's Guide, 0-88175-139-1, $16.00.***

DISCRETE MATHEMATICS: A Computational Approach Using BASIC, ***Marvin Marcus***

"This book is the vanguard of a considerable number of much needed books on discrete mathematics for (college) freshmen and sophomores . . . the intellectual level required of the student would be substantially less than that of a normal freshman calculus course."
—*Computing Reviews*

1983, 0-914894-38-2, 329 pages, $24.95 paper; Apple Diskette, 0-88175-001-8, $20,00; IBM PC Diskette, 0-88175-091-3, $20,00; Solutions Manual, 0-88175-005-0, $10.00.

AN INTRODUCTION TO PASCAL AND PRECALCULUS, ***Marvin Marcus***

This unique new book extends and updates introductory college mathematics courses by integrating the mathematical content with two important developments: inexpensive microcomputers and the Pascal programming language. The mathematical content covers college algebra and elementary functions, using the Apple II Plus or IIe and Pascal. ***1986, 423 pages, 0-88175-009-3, $29.95; Student's Diskette, 0-88175-061-1, $20.00; Teacher's Diskette, 0-88175-062-X, $20.00; Teacher's Solutions Manual, 0-88175-063-8, $10.00.***

PRINCIPLES OF COMPUTER SCIENCE: Concepts, Algorithms, Data Structures, and Applications, ***M. Sandra Carberry, A. Toni Cohen, and Hatem M. Khalil***

Attention is given to each component of computer science in this well-balanced introduction to the field, including algorithm development via stepwise refinement, analysis and verification techniques, components of the Pascal programming language, heuristic problem solving, and more. ***1986, 636 pages, 0-914894-79-X, $34.95; Teacher's Guide with examination questions, 0-88175-154-5, $10.00.***

Complimentary examination copies of college texts are available with a written request on school letterhead. Please identify course name and number, estimated course enrollment, and current text being used.

CSP also has a large selection of computer science titles.
Send for a free catalog or write for further information.

All prices subject to change without notice. Postage and handling is $2.50 for the first book, $1.50 for each additional title. All orders from individuals must be prepaid. All state and local taxes apply.

COMPUTER SCIENCE PRESS, INC. • 1803 Research Blvd., Dept. AMM • Rockville, MD 20850, U.S.A. • (301) 251-9050

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, DC 20036

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 93, Number 9 November 1986

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. (This is necessary for indexing purposes.) Send all main articles, letters to the editor, and editorial correspondence to the editor-elect: HERBERT S. WILF, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Contributions to the other four special departments (Notes, Teaching of Mathematics, Unsolved Problems, and Problems and Solutions) should be sent to the appropriate one of the associate editors, as follows:

Notes: CAROL G. CRAWFORD, Department of Mathematics, U.S. Naval Academy, Annapolis, MD 21402, *or* RICHARD LIBERA, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716, *or* ANITA E. SOLOW, Department of Mathematics, Grinnell College, Grinnell, IA 50112.

Teaching of Mathematics: JOAN P. HUTCHINSON *or* STAN WAGON, Department of Mathematics, Smith College, Northampton, MA 01063.

Unsolved Problems: RICHARD GUY, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.

Proposed problems and solutions, both elementary and advanced: PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801.

Reprint Permission: A. B. WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Back Issues: P. and H. BLISS Co., Middletown, CT 06457.

Microfilm Editions: University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

Editorial Assistants: Fanny Hailpern, Mary P. Jackson.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT, ten times a year: January, February, March, April, May, June-July, August-September, October, November and December.

The annual subscription price for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is $24 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $24. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $70 per year.

Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036

PRINTED IN THE UNITED STATES OF AMERICA

FOURIER'S METHOD OF LINEAR PROGRAMMING AND ITS DUAL

H. P. WILLIAMS
Faculty of Mathematical Studies, University of Southampton, Southampton SO9 5NH, U.K.

Introduction. There has been widespread popular interest in recent years in suggested improved methods for solving Linear Programming (LP) models. In 1977 Shor [**13**] described a new algorithm for LP. Khachian [**7**] modified this algorithm in order to prove that the number of computational steps was, in the worst case, bounded by a polynomial function of the size of the data. This method has become known as the *Ellipsoid Method*. It has in practice been disappointing in experimental computational performance. In 1984 Karmarker [**6**] produced another algorithm which was also "polynomially bounded" with spectacular practical computational claims. Controversy continues as to whether Karmarker's method will displace the Simplex Method. *The Simplex Method* was invented by Dantzig in 1948 and is well explained in Dantzig [**1**]. Although it is not polynomial in the worst case it has proved a remarkably powerful method in practice and its major extension, the *Revised Simplex Method*, is the method used in all commercial systems.

The reason for the widespread popular interest (both Khachian and Karmarker's methods received headlines in the national press) is that LP models are among the most widely used type of Mathematical Model. Applications of LP arise in Manufacturing, Distribution, Finance, Agriculture, Health, Energy and general Resource Planning. A practical discussion of application areas is contained in Williams [**16**].

In this article we show that, predating all these methods, a method discovered by Fourier in 1826 for manipulating linear inequalities can be adapted to Solving Linear Programming models. The theoretical insight given by this method is demonstrated as well as its clear geometrical interpretation. By considering the dual of a linear programming model it is shown how the method gives rise to a dual method. This dual method generates all extreme solutions (including the optimal solution) to a linear programme. Therefore if a polytope is defined in terms of its facets the dual of Fourier's method provides a method of obtaining all vertices.

An LP model consists of variables (e.g., x_1, $x_2, \ldots$, etc.) contained in a linear expression known as an *objective function*. Values are sought for the variables which *maximise* or *minimise* the objective function subject to *constraints*. These constraints are themselves linear expressions which must be either less-than-or-equal to ($\leqslant$), greater-than-or-equal to ($\geqslant$) or equal to ($=$) some specified value. For example, the following is a small LP model.

Find values for $x_1, x_2, \ldots$ among the real numbers so as to:

$$\text{Maximise} \quad -4x_1 + 5x_2 + 3x_3$$

$$\text{P} \quad \begin{array}{c}\text{subject to}\\ \text{constraints}\end{array} \left\{\begin{array}{l} -x_1 + x_2 - x_3 \leqslant 2 \\ \quad x_1 + x_2 + 2x_3 \leqslant 3 \\ x_1 \geqslant 0, x_2 \geqslant 0, x_3 \geqslant 0. \end{array}\right.$$

It is usually the case that the variables are restricted to be non-negative as in the example above.

Paul Williams was born in Cornwall, England, in 1943 and educated at Redruth Grammar School. He graduated in Mathematics from Cambridge University as well as obtaining an Athletics Blue as a hurdler. This was followed by a Ph.D. in Mathematical Logic from Leicester University. It was during this time that he first "discovered" the procedure described in this paper only to find out, some years later, about Fourier's work.

He worked for IBM for a number of years developing Mathematical Programming Software and liaising with clients. In 1976 he was appointed to the first Chair of Management Science at Edinburgh University. Then, in 1984, he moved to the Chair of Operational Research at Southampton University. He is the author of a well-known book "Model Building in Mathematical Programming". His main research interest is in Integer programming.

Professor Williams is married with three children. He is still most at home in Cornwall where he has a cottage and spends as much time as he can.

In practical applications there are sometimes thousands of variables and constraints (a mixture of $\leqslant$, $\geqslant$ and $=$). Typical objective functions represent *profit* (to be maximised) or *cost* (to be minimised).

It is not widely known that in 1826 the French mathematician Fourier [5] devised a method of manipulating Linear Inequalities. He was not concerned with optimising any expression but rather with deriving the *set of solutions* to a system of inequalities (in an analogous way to solving a set of simultaneous equations). His method has been rediscovered a number of times in different contexts. A brief account of some of these is given later.

Fourier's method can comparatively easily be adapted to solving LP models, i.e., *Optimising* an objective function subject to linear inequalities and equations. While the method results in prohibitively large storage requirements for anything but small models it is extremely illuminating and much easier to understand than the Simplex Algorithm. In addition it is a clear way of demonstrating certain theoretical properties of LP models as well as providing more information about other possible solutions.

The method can also be used in a *dual* form to provide another algorithm for solving LP models which generates all *vertex solutions*. Geometrical interpretations of both the original (known as the *primal*) method and the dual method are given later.

Fourier's method. In order to demonstrate Fourier's method we will consider an LP model in a standard form as a maximisation subject to $\leqslant$ constraints. Clearly any model can be converted into this standard form.

When we try to solve an LP one of three possibilities results.

(i) The model is *infeasible*, i.e., there are no values for the variables which satisfy all constraints simultaneously.

(ii) The model is *unbounded*, i.e., the value of the objective function can be increased without limit by choosing values for the variables.

(iii) The model is *solvable*, i.e., there exists a set of values for the variables giving a finite optimal value to the objective function.

Although case (iii) applies to our illustrative numerical example, it will be obvious in the method how cases (i) and (ii) manifest themselves.

In order to demonstrate the method we will use the model P above. Since we wish to maximise $-4x_1 + 5x_2 + 3x_3$ as well as solve the inequalities we will consider the model in the form:

$$\text{P1} \qquad \begin{array}{lrcl}
\text{Maximise } z & & & \\
\text{subject to:} & 4x_1 - 5x_2 - 3x_3 + z & \leqslant 0 & \text{C0} \\
& -x_1 + x_2 - x_3 & \leqslant 2 & \text{C1} \\
& x_1 + x_2 + 2x_3 & \leqslant 3 & \text{C2} \\
& -x_1 & \leqslant 0 & \text{C3} \\
& -x_2 & \leqslant 0 & \text{C4} \\
& -x_3 & \leqslant 0. & \text{C5}
\end{array}$$

Constraint C0 is really a way of saying we wish to maximise z where

$$z \leqslant -4x_1 + 5x_2 + 3x.$$

By maximising z we will "drive" it up to the maximum value of the objective function. It would clearly be possible to treat C0 as an equation but for simplicity of exposition we are treating all constraints as $\leqslant$ inequalities.

Fourier gives a method of *eliminating* variables from inequalities. We will eliminate $x_1, x_2, \ldots,$ etc., from the inequalities $\text{C0}, \text{C1}, \ldots,$ etc., until we are left with inequalities in z above. Then the maximum possible value of z will be apparent.

To eliminate a variable from a set of *inequalities*, Fourier pointed out that we must consider

all *pairs of inequalities* in which the inequality has opposite sign and eliminate between each pair. To demonstrate this we will first consider the import of constraints C0 and C1 above.

C0 can be written as

$$4x_1 \leqslant (5x_2 + 3x_3 - z);$$

C1 can be written as

$$x_1 \geqslant -2 + x_2 - x_3.$$

Therefore we have

$$-2 + x_2 - x_3 \leqslant x_1 \leqslant \frac{1}{4}(5x_2 + 3x_3 - z). \tag{1}$$

Since x_1 is a real number and the real numbers form a continuum (in contrast to the natural numbers), the import of the pair of inequalities above is that

$$-2 + x_2 - x_3 \leqslant \frac{1}{4}(5x_2 + 3x_3 - z), \tag{2}$$

i.e.,

$$-x_2 - 7x_3 + z \leqslant 8. \tag{3}$$

This constraint is more easily arrived at by simply adding 4 times constraint C1 to C0 above to eliminate x_1.

We have shown that if there is a solution to the inequalities such as C0 and C1, there must be a solution to the derived inequality (3). Conversely, if there is a solution to an inequality such as (3), writing it in the form (2) demonstrates that there exists a value of x_1 satisfying (1). In order to give x_1 a value we can take the value of either the left-hand-side or the right-hand-side of the inequality (2) (or any value in between).

Since x_1 also occurs in constraints C2 and C3, we must also eliminate it between all the other pairs in which it has opposite sign, i.e., (C0, C3), (C1, C2) and (C2, C3). If we fail to consider every possible pair we are in danger of losing information and generating spurious (infeasible) solutions.

These eliminations result in the transformed model:

Maximise z

$$\text{P2} \qquad \begin{array}{lrrrl} \text{subject to:} & -x_2 & -7x_3 & +z \leqslant 8 & \text{C0} + 4\text{C1} \\ & -5x_2 & -3x_3 & +z \leqslant 0 & \text{C0} + 4\text{C3} \\ & 2x_2 & + x_3 & \leqslant 5 & \text{C1} + \text{C2} \\ & x_2 & +2x_3 & \leqslant 3 & \text{C2} + \text{C3} \\ & -x_2 & & \leqslant 0 & \text{C4} \\ & -x_3 & & \leqslant 0. & \text{C5} \end{array}$$

The origins of the combined constraints are indicated. It is convenient (but not strictly necessary) always to keep the coefficient of z (the objective), where it occurs, as 1. In order to do this, for example, we add 4 times C1 to C0 in preference to $\frac{1}{4}$ times C0 to C1 or any other combination which reduces the new coefficient of x_1 to zero.

If P1 has a solution (giving values for x_2, x_3 and z), then we have shown P2 must have a solution. Conversely, if P2 has a solution, then a value of x_1 can be found, which satisfies P1, using the argument above.

It is worth contrasting this elimination procedure (for inequalities) with Gaussian elimination (for equations). If the variable to be eliminated has a nonzero coefficient in an equation, this equation (known as the pivot equation) can be used to eliminate the variable from all other equations. With inequalities our elimination procedure is clearly more complex, although should such an equation be present (together with inequalities) we can still use it in this way.

Having eliminated x_1 we now eliminate another variable. There is complete flexibility in the

order in which the variables are eliminated. For convenience we will continue to eliminate the variables in consecutive order and to choose x_2. The pairs of constraints in which x_2 have opposite sign are

$$(C0 + 4C1, C1 + C2), (C0 + 4C3, C1 + C2), (C1 + C2, C4),$$
$$(C0 + 4C1, C2 + C3), (C0 + 4C3, C2 + C3) \quad \text{and} \quad (C2 + C3, C4).$$

Combining those constraints in suitable multiples in order to eliminate x_2 reduces the model to:

$$\text{P3} \qquad \begin{array}{llll} \text{Maximise } z & & & \\ \text{subject to:} & -\frac{13}{2}x_3 + z \leqslant \frac{21}{2} & & (C0 + 4C1) + \frac{1}{2}(C1 + C2) \\ & -\frac{1}{2}x_3 + z \leqslant \frac{25}{2} & & (C0 + 4C3) + \frac{5}{2}(C1 + C2) \\ & x_3 \leqslant 5 & & (C1 + C2) + 2C4 \\ & -5x_3 + z \leqslant 11 & & (C0 + 4C1) + (C2 + C3) \\ & 7x_3 + z \leqslant 15 & & (C0 + 4C3) + 5(C2 + C3) \\ & 2x_3 \leqslant 3 & & (C2 + C3) + C4 \\ & -x_3 \leqslant 0. & & C5 \end{array}$$

It has been shown by Kohler [**8**] that after n variables have been eliminated any constraint that depends on more than $n + 1$ of the original constraints must be redundant (implied by the other constraints). In this case after eliminating 2 variables the 2nd and 4th of the above inequalities depend on more than 3 of the original inequalities (both depend on C0, C1, C2 and C3). Therefore Kohler's result allows us to ignore the 2nd and 4th inequalities giving the representation:

$$\text{P3}' \qquad \begin{array}{lll} \text{Maximise } z & & \\ \text{subject to:} & -\frac{13}{2}x_3 + z \leqslant \frac{21}{2} & C0 + \frac{9}{2}C1 + \frac{1}{2}C2 \\ & x_3 \leqslant 5 & C1 + C2 + 2C4 \\ & 7x_3 + z \leqslant 15 & C0 + 5C2 + 9C3 \\ & 2x_3 \leqslant 3 & C2 + C3 + C4 \\ & -x_3 \leqslant 0. & C5 \end{array}$$

Finally we eliminate x_3 between pairs of inequalities where x_3 has coefficients of opposite sign. Again Kohler's result enables us to ignore the elimination between the 1st and 4th constraint. The resultant transformed model is:

$$\text{P4} \qquad \begin{array}{lll} \text{Maximise } z & & \\ \text{subject to:} & z \leqslant 43 & C0 + 11C1 + 7C2 + 13C4 \\ & 0 \leqslant 5 & C1 + C2 + 2C4 + C5 \\ & z \leqslant \frac{38}{3} & C0 + \frac{7}{3}C1 + \frac{8}{3}C2 + \frac{13}{3}C3 \\ & z \leqslant 15 & C0 + 5C2 + 9C3 + 7C5 \\ & 0 \leqslant 3. & C2 + C3 + C4 + 2C5 \end{array}$$

Clearly the maximum value of z satisfying *all* these constraints is $38/3$. This arises as the *minimum* constant on the right-hand side of the three inequalities involving z. In order to obtain the values of the variables $x_1, x_2, \ldots,$ etc., which give rise to the maximum value of z we can work backwards as follows.

The 3rd constraint in P4 is that one which gives $z = 38/3$. This arises from combining the 1st and 3rd constraints in P3′. If $z = 38/3$ (instead of $z < 38/3$), then we must have the 1st and 3rd

constraints of P3′ satisfied as equations. Solving these equations gives $x_3 = 1/3$. These constraints in turn arise from the 1st, 2nd, 3rd and 4th constraints in P2 which when solved as equations give $x_2 = 2\frac{1}{3}$. Finally the origins of these constraints are C0, C1, C2, and C3 in P1 which when solved as equations give $x_1 = 0$.

Alternatively we could observe immediately that constraint $z \leqslant 38/3$ in P4 arises from C0, C1, C2 and C3. If we set $z = 38/3$, this forces us to treat these constraints as equations, which when solved simultaneously give this optimal solution.

This method gives us much more information that just the specific solution to a specified model. The coefficients (multipliers) of C0, C1, C2 and C3 in the 3rd constraint of P4 are 1, 7/3, 8/3 and 13/3, since C0 in P1 consists of the negated original objective (plus z) this points out the obvious result that

$$\begin{aligned}
&\frac{7}{3}(-x_1 + x_2 - x_3 \leqslant 2)\\
+&\frac{8}{3}(x_1 + x_2 + 2x_3 \leqslant 3)\\
+&\frac{13}{3}(-x_1 \qquad\qquad \leqslant 0)\\
&\qquad\qquad\downarrow\\
&-4x_1 + 5x_2 + 3x_3 \leqslant \frac{38}{3}
\end{aligned}$$

These multipliers therefore show 38/3 to be an *upper bound* for the maximum value of $-4x_1 + 5x_2 + 3x_3$.

Similarly the multipliers of $C_1, C_2, \dots$, etc., in the other inequalities (the 1st and 4th) in P4 involving z give (non-strict) upper bounds of 43 and 15, respectively, for the objective.

Our method has not only provided us with multipliers for the constraints and an upper bound for the objective function. It has also provided us with a set of values for the variables for which the objective attains the least upper bound derived. This is the main import of the famous *Duality Theorem* of LP which is discussed later.

The significance of the other inequalities $0 \leqslant 5$ and $0 \leqslant 3$, not involving z, in P4 will become apparent when we describe the dual of the method above.

Should an LP model be *infeasible* the method demonstrates this. The final inequalities will contain a *contradiction*, i.e., a constraint such as

$$0 \leqslant -1.$$

If a model is *unbounded*, this will be apparent as in the final inequalities there will be no upper limit to the value of z.

Although we have solved model P for specific values of the right-hand-side coefficients of the inequalities, it should be apparent that those values were not used until we derived the maximum value of z from P4. Therefore we could, with no extra work, have found the maximum value of z as a function of the right-hand-side coefficients. Such a function is known as the *value function* of an LP. If the right-hand-side values of the two constraints (apart from the nonnegativity constraints) in P were b_1 and b_2 instead of 2 and 3, the multipliers of C0, C1, etc., in P4 would tell us that the final inequalities would be

$$\begin{aligned}
z &\leqslant 11b_1 + 7b_2\\
0 &\leqslant \ b_1 + b_1\\
z &\leqslant \frac{7}{3}b_1 + \frac{8}{3}b_2\\
z &\leqslant 5b_1 + 9b_2\\
0 &\leqslant \ b_2.
\end{aligned}$$

Therefore if $b_1 + b_2$ or b_2 is negative, the model is infeasible. Otherwise

$$z = \text{minimum}\left(11b_1 + 7b_2, \frac{7}{3}b_1 + \frac{8}{3}b_2, 5b_1 + 9b_2\right).$$

A geometrical interpretation of the method. The method can be interpreted geometrically. It is possible to represent the model P1 in 4-dimensional Euclidean space. A point represents a *feasible* solution if its coordinates give values for the variables which satisfy the constraints. The set of feasible solutions can be shown to give a *polyhedron* in 4 dimensions. For a general model with n variables we will have a *polyhedron* in n dimensions. (The polyhedron may not be bounded as in this example.) By eliminating a variable we *project* the polyhedron down into a space of one less dimension. While we cannot visualise a space of 4 dimensions, we can visualise the transformed model P2 which has been reduced to 3 variables, and therefore is represented in a space of 3 dimensions in Fig. 1(i). By maximising z we are trying to find the *highest* point in this three-dimensional polyhedron. Each of the inequalities in P2 gives rise to a 2-dimensional face of the polyhedron. These are the faces ABC, $ACED$, $ABHI$, $ADFI$, $FDEG$ and $GECBH$. In order to visualise the diagram more easily, the coordinates (x_2, x_3, z) of the 5 vertices A, B, C, D and E are marked. The lines FD, GE, HB and IA are all parallel to the z axis. In this example none of the inequalities is redundant. If there were redundant inequalities, these would give rise to 2-dimensional planes outside the polyhedron and not therefore forming boundaries.

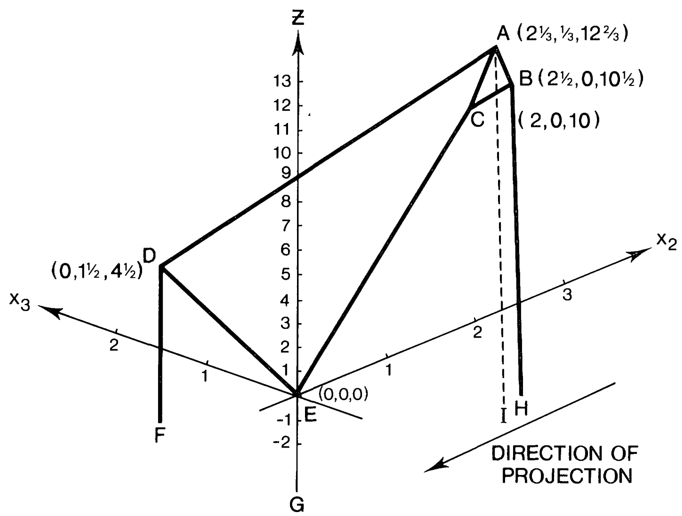


FIG. 1(i)

The elimination of variable x_2 projects this polyhedron down onto the plane (x_3, z) giving the model P3 (or P3′). In effect what we are doing by eliminating x_2 is shining rays of light parallel to the axis x_2 in the direction of the (x_3, z) plane. The *shadow* of the 3-dimensional polyhedron on the plane gives the polyhedron associated with P3′ represented in Fig. 1(ii). The inequalities in P3′ (apart from the second) respectively give rise to the lines PQ, PR, RT and QS. Although Kohler's observation allowed us to remove some redundant inequalities in P to produce $P3'$, it does not remove them all. From Fig. 1(i) it is apparent that the inequality $x_3 \leqslant 5$ is redundant (implied by the other inequalities). PQ, PR, RT and QS form the 1-dimensional faces of the 2-dimensional polyhedron. (In fact the inequality $x_3 \leqslant 5$ is the "shadow" of the line of intersection of the extended faces $ABHI$ and $FDEG$ in Fig. 1(i).)

Finally, eliminating x_3, we project the 2-dimensional polyhedron in 1(ii) down onto the z axis

to give the 1-dimensional polyhedron in Fig. 1(iii). The 1st (redundant) inequality in P4 gives the point $z = 43$ which is not marked. The 3rd inequality gives the point X and the 4th inequality the point Y. Clearly X is the only 0-dimensional face of the 1-dimensional polyhedron and all inequalities apart from the 4th are redundant. For completeness we observe that the point $z = 43$ is the shadow of the intersection of lines QP and $x_3 = 5$ in Fig. 1(ii). Point Y is the shadow of the intersection of the extensions of SQ and PR. The redundant inequality $0 \leqslant 5$ is the "shadow" of the "intersection" of the parallel lines SQ and $x_3 = 5$; similarly $0 \leqslant 3$ is the "shadow" of the "intersection" of RT and SQ.

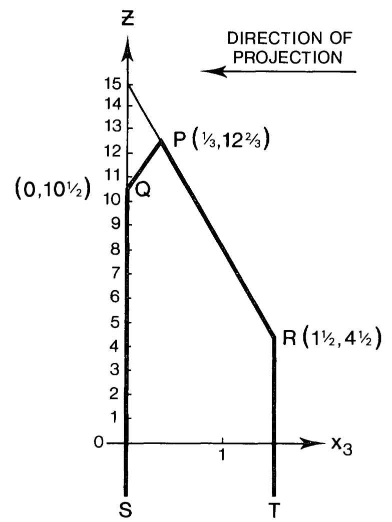


FIG. 1(ii)

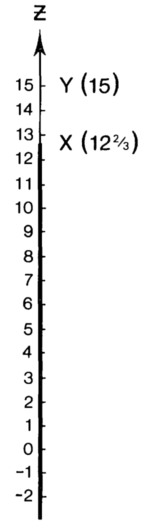


FIG. 1(iii)

Having shown that the maximum possible value of z arises from point X in 1(iii), we backtrack to the point P in 1(ii) of which X is the shadow giving $x_3 = 1/3$. P is the shadow of A in 1(i) giving $x_2 = 2\frac{1}{3}$. If it were possible to visualise 4 dimensions, A would be the shadow of a vertex of the 4-dimensional polyhedron represented by P1.

Were the original model to be *infeasible*, it would be represented by an empty polyhedron whose projections would clearly be empty. If the model were unbounded, the polyhedron would be unbounded in the z-direction which would be revealed in the projection onto the z axis.

In practice the build-up in inequalities resulting from the elimination of each variable can be explosive. If, for example, a variable to be eliminated occurs with a negative coefficient in m_1 inequalities, a positive coefficient in m_2 inequalities, and does not occur in the remaining m_3 inequalities, the result of eliminating it will be to produce $m_1 m_2 + m_3$ inequalities. Many of these resultant inequalities will be redundant. Although Kohler's observation may allow us to remove some of them, the number can still become very large even for quite modest values of m_1 and m_2. It is this, potentially explosive, growth in inequalities which makes the method computationally impractical for real life models. No efficient method has yet been devised for removing all the redundant inequalities generated.

The dual model. Another illuminating way of looking at the method is to consider the *dual model*. It has already been pointed out that the multipliers of $C0, C1, C2, \dots$, etc., in P4 demonstrate different ways in which the constraints of P can be added together to give an upper bound for the objective function. If we look at the rows of *detached coefficients* of x_1, x_2 and x_3 in the constraints $C1, C2, \dots,$ we have

−1	1	−1	C1
1	1	2	C2
−1	0	0	C3
0	−1	0	C4
0	0	−1	C5

The multipliers of C1, C2, ..., etc., (which will always be nonnegative) in the 1st, 3rd and 4th inequalities of P4, give different ways in which these rows can be added together to give the rows of detached coefficients of the objective in P i.e.,

−4	5	3

If we let the multipliers of C1, C2, ..., C5 be $y1$, $y2$, ..., $y5$, we must have

$$\begin{aligned} -y_1 + y_2 - y_3 \qquad &= -4 \\ y_1 + y_2 \qquad - y_4 \quad &= 5 \\ -y_1 + 2y_2 \qquad\qquad - y_5 &= 3 \end{aligned}$$

The multipliers for the 1st inequality in P4 provide a solution to this set of equations

$$y_1 = 11, \quad y_2 = 7 \quad y_3 = 0, \quad y_4 = 13, \quad y_5 = 0.$$

The multipliers for the 3rd inequality in P4 provide another solution to the equations

$$y_1 = \frac{7}{3}, \quad y_2 = \frac{8}{3}, \quad y_3 = \frac{13}{3}, \quad y_4 = y_5 = 0.$$

The multipliers for the 4th inequality in P4 provide yet another solution to the equations.

$$y_1 = 0, \quad y_2 = 5, \quad y_3 = 9, \quad y_4 = 0, \quad y_5 = 7.$$

What we are seeking are a set of nonnegative multipliers (values for the y variables) which give the *least upper bound* for the objective, in P. In order to do this we wish to

$$\text{Minimise} \quad 2y_1 + 3y_2.$$

where these coefficients 2 and 3 are the values on the right-hand sides of C1 and C2 in P1.

The problem which we have posed involving variables $y_1, y_2, \ldots$, etc., is itself an LP model. The variables y_3, y_4 and y_5 in the above three equations are sometimes known as *surplus variables*. Since they (like all the variables) cannot take negative values, the three equations above can be written as $\geqslant$ inequality constraints. If the expression $2y_1 + 3y_2$ is regarded as the new objective function, we have the new model in the form:

$$\text{D} \qquad \begin{aligned} &\text{Minimise} && 2y_1 + 3y_2 \\ &\text{subject to:} && -y_1 + y_2 \geqslant -4 \\ & && y_1 + y_2 \geqslant 5 \\ & && -y_1 + 2y_2 \geqslant 3 \\ & && y_1 \geqslant 0, \quad y_2 \geqslant 0. \end{aligned}$$

This model D is known as the *dual model* to the (*primal*) model P.

We have already, indirectly, found, by Fourier's method, a solution to D, where the value of the objective is equal to the maximum possible value of the objective of P. Since it should now be apparent that any solution to D provides an upper bound for the maximum objective of P, the solution we have obtained for D must also minimise the objective of D. This, it has already been pointed out, is an instance of a general powerful and famous result known as the *Duality Theorem* of LP. To every LP model there corresponds a dual model. If both are solvable (i.e., not infeasible or unbounded) the optimal objective values of both are the same. Fourier's method provides a clear demonstration of this.

The dual method. The fact that every LP model has a dual model allows us to convert Fourier's method into a *dual method*. Each of the steps in our original (primal) method applied to the original model can be mirrored by steps applied to the dual model. The resultant method is also intuitive and has a clear geometrical interpretation.

In our primal method we combined *rows* (constraints) together, two at a time, so as to eliminate variables (columns) from the model. Ultimately we arrived at nonnegative combinations of the original rows which gave the objective function. For the dual method we will combine *columns* together, two at a time, so as to eliminate constraints (rows) from the model. Ultimately we will arrive at nonnegative combinations of the columns which give the column of right-hand-side coefficients of the model. The multipliers in these non-negative linear combinations will constitute feasible solutions to the dual model. We seek a feasible solution which minimises the dual objective function.

Just as it was convenient to convert model P into model P1 by representing the objective by a variable z, it is convenient in the dual method, applied to model D, to represent the right-hand-side constants as coefficients of a new variable y_0 fixed at value 1. We also, in P1, explicitly included the nonnegativity conditions $-x_1 \leqslant 0$, etc. The dual correspondence to this is to include the surplus variables so making the constraints of D into equations. This gives us the form D1 of the model.

$$
\text{D1}\qquad
\begin{array}{llrl}
\text{Minimise} & 2y_1 + 3y_2 & & \\
\text{subject to:} & 4y_0 - y_1 + y_2 - y_3 & = 0 & \text{A1} \\
 & -5y_0 + y_1 + y_2 - y_4 & = 0 & \text{A2} \\
 & -3y_0 - y_1 + 2y_2 + y_5 & = 0 & \text{A3} \\
 & y_0 & = 1 & \text{B} \\
 & y_1, y_2, y_3, y_4, y_5 \geqslant 0. & &
\end{array}
$$

The parallel between P1 and D1 should be obvious. The coefficients in the four *rows* of D1 are the same as the coefficients in the four *columns* of P1. The objective coefficients of D1 are the same as the right-hand-side coefficients of P1.

In order to eliminate constraint A1 we apply a transformation of variables. New (nonnegative) variables u_1, u_2, u_3 and u_4 are introduced which are related to the variables $y_0, y_1, \ldots,$ etc., by the equations

$$
\begin{aligned}
4u_1 + u_3 &= y_1, & 4y_2 + u_4 &= y_3, \\
4u_1 + 4u_2 &= 4y_0, & u_3 + u_4 &= y_2.
\end{aligned}
$$

When these equations are used to substitute y_0, y_1, y_2 and y_3 out of the equations in D1 it can easily be verified that the equation A1 disappears.

A graphic way of interpreting the row variables has been suggested by Dantzig and Eaves [**2**].This can best be understood through Fig. 2. In equation A1 we have a mixture of negative quantities ($-y_1$ and $-y_3$) and positive quantities ($4y_0$, y_2) which must sum to zero to satisfy the equation. $4y_0$ is split up into $4u_1$ and $4u_2$ (the coefficients of u_1 and u_2 are kept the same as y_0 so as to keep all the coefficients in the transformed equation $y_0 = 1$ as unity. This is for

convenience rather than necessity). Similarly the other quantities in equation A1 are split up as indicated in Fig. 2. Fig. 2 may be interpreted as a "Transportation Problem." The Transportation problem is itself a particularly simple type of the LP model and is described in Dantzig [**1**].

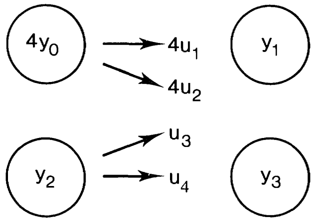


FIG. 2

While visualising the transformed variables in this way gives an interpretation to the new variables, it is not necessary for the execution of the dual method. This can be carried out mechanically by analogy with the primal method, as will become clear through the example. Performing the above substitutions transforms D1 to the model below.

$$\text{D2}\qquad \begin{array}{lrcl} \text{Minimise} & 8u_1 + 5u_3 + 3u_4 & & \\ \text{subject to:} & -u_1 - 5u_2 + 2u_3 + u_4 - y_4 & = 0 & \text{A2} \\ & 7u_1 - 3u_2 + u_3 + 2u_4 - y_5 & = 0 & \text{A3} \\ & u_1 + u_2 & = 1 & \text{B} \\ & u_1, u_2, u_3, u_4, y_4, y_5 \geqslant 0. & & \end{array}$$

It is clear that D2 is the dual of model P2. In this dual method we have removed *constraint* A1 whereas in the primal method we removed *variable* x_1.

Rather than think in terms of transformed variables we can perform the method computationally by combining columns in pairs. The column of coefficients for u_1 in D2 arises as the column for y_0 in D1 added to 4 times the column for y_1. These two columns are combined in these multiples in order to eliminate the coefficient of the new variable u_1 in D2. Similarly the columns for the pairs of variables (y_0, y_3), (y_1, y_2) and (y_2, y_3), each having opposite signs in A1, are combined in suitable multiples. The correspondence with the elimination of x_1 in P1 in the primal method should be apparent. It is convenient to remember the origins of each column. This may conveniently be done by Table 1 below.

TABLE 1

	u_1	u_2	u_3	u_4	y_4	y_5
y_0	1	1				
y_1	4		1			
y_2			1	1		
y_3		4		1		
y_4					1	
y_5						1

The elimination of A2 from D2 can be performed similarly by combining columns for the pairs (u_1, u_3), (u_2, u_3), (u_3, y_4), (u_1, u_4), (u_2, u_4) and (u_4, y_4) in suitable multiples so that the resultant coefficients in A2 are all zero. As with the primal method some of these combinations can be ignored. If after n constraints have been eliminated a column depends upon more than $n + 1$ of the original columns, it can be shown that it may be ignored. This is the obvious dual of Kohler's observation in the primal method. The reason why it is possible to ignore such columns is pointed out below. For our example here it means that we need not combine the pairs (u_2, u_3)

and (u_1, u_4). Both would result in columns depending on y_0, y_1, y_2 and y_3. The result of eliminating constraint A2 is to produce D3′ (the dual of P3′).

$$\text{D3}' \qquad \begin{array}{llr} \text{Minimise} & \frac{21}{2}v_1 + 5v_2 + 15v_3 + 3v_4 & \\ \text{subject to:} & -\frac{13}{2}v_1 + v_2 + 7v_3 + 2v_4 - y_5 = 0 & \text{A3} \\ & v_1 + v_3 = 0 & \text{B} \\ & v_1, v_2, v_3, v_4, y_5 \geqslant 0. & \end{array}$$

The origins of the columns for the variables are given in Table 2.

TABLE 2

	v_1	v_2	v_3	v_4	y_5
y_0	1		1		
y_1	$\frac{9}{2}$	1			
y_2	$\frac{1}{2}$	1	5	1	
y_3			9	1	
y_4		2		1	
y_5					1

Table 2 can be constructed by combining the columns of Table 1 in the same multiples as the columns of D2. For example, the column for v_1 in $D3'$ arises from the column for u_1 in D2 added to $\frac{1}{2}$ times the column for u_3 in D2. Similarly, the column for v_1 in Table 2 is the column for u_1 in Table 1 added to $\frac{1}{2}$ times the column for u_3 in Table 1. Multiples of columns are chosen so as to keep the nonzero coefficients of y_0 unity in the tables of originating variables.

Finally, eliminating A3 from D3′ produces model D4 and Table 3.

$$\text{D4} \qquad \begin{array}{llr} \text{Minimise} & 43w_1 + 5w_2 + \frac{38}{3}w_3 + 15w_4 + 3w_5 & \\ \text{subject to:} & w_1 + w_3 + w_4 = 1 & \text{B} \\ & w_1, w_2, w_3, w_4, w_5 \geqslant 0. & \end{array}$$

TABLE 3

	w_1	w_2	w_3	w_4	w_5
y_0	1		1	1	
y_1	11	1	$\frac{7}{3}$		
y_2	7	1	$\frac{8}{3}$	5	1
y_3			$\frac{13}{3}$	9	1
y_4	13	2			1
y_5		1		7	2

The solution of D4 is obvious. We choose that variable from among w_1, w_3 and w_4 which has the smallest objective coefficient and set it to 1. Clearly this gives $w_3 = 1$. From Table 3 we see the multiples of the original columns of the model D1 which give rise to the column for w_3 in D4. Therefore the optimal solution to the original model is

$$y_1 = \frac{7}{3}, \quad y_2 = \frac{8}{3}, \quad y_3 = \frac{13}{3}, \quad y_4 = y_5 = 0$$

given an objective value of 38/3. The coefficient 1 for row y_0 in Table 3 indicates that we must take 1 times the negated column of right-hand-side coefficients for 0 in making up the optimal solution.

It should be obvious that again D4 is the dual model to P4 and contains the same coefficients but these are transposed. The coefficients in Table 3 are the same as the multipliers of the original constraints of P given in P4. In the same way that the primal method can provide the optimal solution for any right-hand-side coefficient, this dual method gives the optimal solution for *any objective function*. If model D were to have another objective function, then the final transformed model D4 would be the same apart from its objective coefficients. In fact, if the objective coefficients in D were b_1 and b_2 instead of 2 and 3, the transformed model D4 would be:

$$\begin{aligned} &\text{Minimise} \quad (11b_1 + 7b_2)w_1 + (b_1 + b_2)w_2 + \left(\frac{7}{3}b_1 + \frac{8}{3}b_2\right)w_3 + (5b_1 + 9b_2)w_4 + b_2 w_5 \\ &\text{subject to:} \quad w_1 + w_3 + w_4 = 1 \\ &\qquad w_1, w_2, w_3, w_4, w_5 \geqslant 0. \end{aligned}$$

If $b_1 + b_2$ or b_2 is negative, the objective can be made as small as we like and the model is said to be *unbounded* (the primal model was infeasible in these cases); otherwise the minimum value of the objective is

$$\text{minimum}\left(11b_1 + 7b_2, \frac{7}{3}b_1 + \frac{8}{3}b_2, 5b_1 + 9b_2\right).$$

The satisfaction of the duality theorem should again be obvious. The corresponding values of the variables $y_1, y_2, \ldots,$ etc., are given in the corresponding column of Table 3. Therefore apart from the case of the model D being unbounded there are three possible optimal solutions. They are:

$$\begin{aligned} &\text{corresponding to } w_1 = 1: y_1 = 11, \quad y_2 = 7, \quad y_3 = 0, \quad y_4 = 13, \quad y_5 = 0, \\ &\text{corresponding to } w_3 = 1: y_1 = \frac{7}{3}, \quad y_2 = \frac{8}{3}, \quad y_3 = \frac{13}{3}, \quad y_4 = y_5 = 0, \\ &\text{corresponding to } w_4 = 1: y_1 = 0, \quad y_2 = 5, \quad y_3 = 9, \quad y_4 = 0, \quad y_5 = 7. \end{aligned}$$

These three solutions are obviously the three sets of multipliers for the constraints on the final form of the primal model P4. In the dual model D they are three *vertex solutions*. Model D is represented in Fig. 3. The three constraints of D are represented by the faces CE, AB and BC, respectively. AD represents the nonnegativity constraint on y_1. The nonnegativity constraint on y_2 is clearly redundant. Different objective functions will give either an unbounded solution or one of the three vertex solutions at A, B or C. For example, different values of the objective function $2y_1 + 3y_2$ give lines parallel to PQ. By minimising this objective function we move to the lowest such line which still intersects the feasible region, in this case at vertex B, giving the solution $y_1 = 7/3, y_2 = 8/3$, objective $= 38/3$ already obtained. The lines AD and CE are known as *extreme rays*. Their existence is demonstrated algebraically by the columns for w_2 and w_5 in Table 3 which have entries of 0 in row y_0. For example, we can let w_2 take any nonnegative value without violating constraint B of D4. This corresponds to keeping y_1 and y_2 in the ratio 1:1 (coefficients in Table 3) and fixing y_3 at 0 (the constraint represented by CE is therefore binding). Clearly the column for w_2 in Table 3 corresponds to the extreme ray CE. Similarly the column for w_5 corresponds to the extreme ray AD.

We have therefore demonstrated that the dual of Fourier's method generates *all vertices and extreme rays* for the feasible polyhedron of an LP model. This in itself sometimes has practical application.

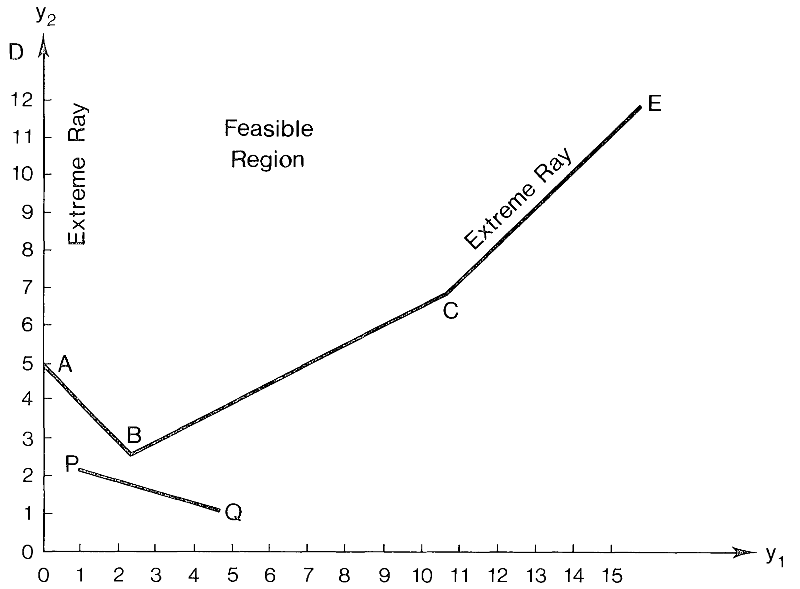


FIG. 3

It is well known that, for an LP model with m constraints, we can restrict our search for an optimal solution to solutions in which at most m variables (including slack and surplus variables) are nonzero. This is an algebraic realisation of the geometric observation that the optimal solution to an LP (if it exists) lies on the boundary of the polyhedron defined by the feasible region. If the optimal solution is unique, it will lie at a vertex, otherwise, in the case of alternate optimal solutions, there will still be among these alternatives vertex solutions which are optimal. The Simplex Algorithm restricts attention to so-called *basic solutions* which correspond to vertex solutions. This property allows us to justify Kohler's observation when applied to the dual method. When we have eliminated n constraints from our original model (D, say) we have in effect solved an LP model consisting of the first n constraints. In the optimal solution to such a model no more than n of the original variables will be nonzero. Therefore, including our right-hand-side column y_0 as a variable, no more than $n + 1$ of the original variables will go to make up a vertex solution. Hence any derived column depending on more than $n + 1$ of the original variables will correspond to a variable which can be taken as 0 in an optimal solution. Hence such a derived column may be ignored. Because of the one-to-one correspondence between derived columns in the dual method and derived rows in the primal method, this is a sufficient justification for our ignoring certain derived constraints. We did this when they depended on more than $n + 1$ of the original constraints when n constraints had been eliminated (Kohler's observation).

An outline of the history of Fourier's method and its extensions. Fourier's method was published 1826. It has been rediscovered a number of times by different authors. Motzkin [**12**] derived a method of solving 2-person zero sum games. Since any LP can be formulated as such a game (and vice versa), Motzkin's method gives rise to a method of solving LP models which in fact turns out to be Fourier's method. Hence the name Fourier-Motzkin elimination is often used for the method. Dantzig [**1**] refers to the method briefly under this name. Dines [**3**] also rediscovered the method. Langford [**10**] derived a method of solving a particular problem in Mathematical Logic. He showed, by a constructive method, that the Theory of Dense Linear Order is decidable. Williams [**14**] showed that any LP model can be posed within this restricted form of arithmetic and that hence the achievability, or otherwise, of a particular objective value

can be decided. This application of Langford's method turns out to be the same as Fourier's method. Another account of Fourier's method, together with additional references, can be found in Duffin [**4**]. There is also a related article by Kuhn [**9**].

Fourier's method (and its dual) is computationally impractical for anything but small models. This is because of the large build-up in inequalities (or variables) as variables (or constraints) are eliminated. It is, however, possible that the methods could be applied in a restricted form. When all variables (apart from the objective variable) have been eliminated, one will only be interested in one of the derived inequalities. For the dual method one will only be interested in one of the final columns. Unfortunately, it is not clear how to eliminate most of the redundant inequalities (or variables) until the end. Williams [**18**] suggests applying a restricted form of the dual method as a "Crashing Procedure" prior to the Simplex Algorithm. Geometrically the Simplex Algorithm moves from vertex solution to vertex solution until it reaches the optimal vertex solution. Initially (Phase 1 of the Simplex Algorithm) it is necessary to obtain a feasible vertex solution. In practice this usually takes as much time as the second phase. For model D represented in Fig. 3 the Simplex Algorithm would start at the origin 0 and systematically move to a vertex (such as A) before proceeding to the optimal vertex at B. By applying a restricted form of the dual method one would hope to obtain a good vertex solution as a starting point.

Computational implementations of the methods using efficient data structures are possible. It is sensible to take account of the sparseness of most LP models (most coefficients in a model are usually zero) in both storing and manipulating the matrices. The transformations which eliminate variables or constraints can be represented by elementary matrices which probably gives a sparser representation than explicitly transforming the whole model. Such considerations are, however, beyond the scope of this paper.

There is a lot of interest, in view of its wide applicability, in an extension of LP known as Integer Programming (IP). Here some, or all, of the variables in a model are restricted to take integer values. Such models are much more difficult to solve than LPs. It has been shown by Lee [**11**] and Williams [**15**] how Fourier's method can be extended to allow us to eliminate integer variables. In order to do this it is necessary to introduce disjunctions of inequalities as well as congruence relations into the transformed model. The dual method can also be extended to deal with IP models by introducing congruence relations as is done by Williams [**17**].

Acknowledgement. The author would like to acknowledge the help of Mr. Robin E. Day of Edinburgh University who wrote a sophisticated computer program to perform the dual method.

References

1. G. B. Dantzig, Linear Programming and Extensions, Princeton, 1963.

2. G. B. Dantzig and B. C. Eaves, Fourier-Motzkin elimination and its dual, J. Combin. Theory Ser. A, 14 (1973) 288–297.

3. L. L. Dines, Systems of linear inequalities, Ann. of Math., 20 (1919) 191–199.

4. R. J. Duffin, On Fourier's Analysis of Linear Inequality Systems, Mathematical Programming Study I, 1974, pp. 71–95.

5. J. B. J. Fourier, Solution d'une Question Particulière du Calcul des Inégalités, Oeuvres II, Paris, 1826, pp. 317–328.

6. N. Karmarkar, A New Polynomial-Time Algorithm for Linear Programming, Proceedings of the 16th annual ACM Symposium of the Theory of Computing, Washington, DC, April 30–May 2, 1984 pp. 302–310.

7. L. G. Khachian, A polynomial algorithm in linear programming, Dokl. Akad. Nauk SSSR, 244 (1979) 1093–96 (translated as Soviet Mathematics Doklady, 20, 191–194).

8. D. A. Kohler, Projections of Convex Polyhedral Sets, Operational Research Center Report, ORC 67–29, University of California, Berkeley, 1967.

9. H. W. Kuhn, Solvability and consistency for linear equations and inequalities, this MONTHLY, 43(1956) 217–232.

10. C. H. Langford, Some theorems on deducibility, Ann. of Math. I, 28 (1927) 16–40.

11. R. D. Lee, An application of mathematical logic to the integer linear programming problem, Notre Dame J. Formal Logic, 23 (1972) 279–282.

12. T. S. Motzkin, Beitrage zur Theorie der Linearen Ungleichungen, Dissertation, University of Basel, Jerusalem, 1936.

13. N. Z. Shor, Cut-off method with space extension in convex programming problems, Kibernetika, 13 (1977) 94–96.

14. H. P. Williams, An Algorithm for the Solution of Linear Programming Problems, IBM Working Paper, 1970.

15. ______, Fourier-Motzkin elimination extension to integer programming, J. Combin. Theory Ser. A, 21 (1976) 118–123.

16. ______, Model Building in Mathematical Programming, Wiley, New York, 1985.

17. ______, A characterization of all feasible solutions to an integer programme, Discrete Applied Math., 5 (1983) 147–155.

18. ______, Restricted vertex generation applied as a crashing procedure for linear programming, Computers and Operations Research, 11 (1984) 401–407.

MUSICAL SCALES AND THE GENERALIZED CIRCLE OF FIFTHS

JOHN CLOUGH
Department of Music, SUNY at Buffalo, Amherst, NY 14260

GERALD MYERSON
Department of Mathematics, The University of Texas at Austin, Austin, TX 78712

This paper deals with the way the diatonic set (the white keys on the piano) is embedded in the chromatic scale (all the keys on the piano). To illustrate the problem, consider the chords CDF and EFA (the reader who happens to be temporarily without piano may find Fig. 1 helpful). If we ignore the black keys, these chords have the same structure; the second note is one key higher than the first, and the third note is two keys higher than the second. When actually played on the piano, the chords sound quite different, due to the embedding of the diatonic in the chromatic. From C to D is two semitones (a semitone is the distance between adjacent notes in the chromatic scale), and from D to F is three, whereas E to F is one and F to A is four. The problem

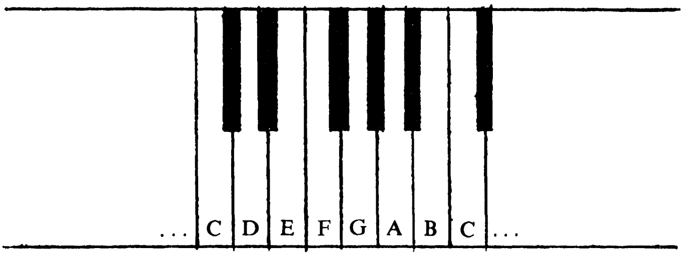


FIG. 1. Piano keyboard.

John Clough: Before coming to SUNY at Buffalo, I taught at the Oberlin College Conservatory of Music and in the School of Music at the University of Michigan. At all three places I have enjoyed the colleagueship of mathematicians who were willing to help me work through various problems in the application of mathematics to music: Edward Wong and Samuel Goldberg at Oberlin, Bernard Galler at Michigan, John Myhill and Gerald Myerson at Buffalo. Though trained only as a musician, in occasional flights of fancy I consider a second career in my first love—mathematics.

Gerald Myerson: I received my Ph.D. in Mathematics under the direction of Don Lewis at the University of Michigan in 1977. I have been on the faculty at the University of Buffalo, the University of British Columbia, and the University of Texas. I play two musical instruments: the phonograph and the cassette deck.

we pose here is, among all the chords of a given structure, how many "different" chords are there? (We remark that a different problem in the enumeration of chords was addressed recently in this MONTHLY [5].)

We first make our language more precise. The white-key set $\{A, B, C, D, E, F, G\}$ will be referred to as the *diatonic set*. (In fact, it is one of many diatonic sets similarly embedded in the chromatic scale; the results we obtain for the white-key set apply to all diatonic sets.) Each element of this set is a class of notes; thus, "C", for example, represents middle C, high C, low C, etc. A *chord* is a non-empty subset of the diatonic set. Juxtaposition will generally be used in preference to set notation; thus, CDF rather than $\{C, D, F\}$. A *line* is a finite non-empty sequence of distinct elements of the diatonic set. Hyphenation will be used for lines; thus, C–D–F is a line. The distinction between lines and chords is that between permutations and combinations; the chords CDF and DCF are identical, the lines C–D–F and D–C–F are distinct. Both concepts are significant in music theory. We will restrict our attention to lines, for which the mathematics is somewhat neater; we leave it to the reader to work out the theory of chords.

An *interval* is a two-note line. The *diatonic length* of an interval is the number of steps (in the diatonic) required to go from the first note of the line to the second—here, and always, we go "up" the scale. For example, the diatonic length of C–F is 3; that of F–C is 4. Notice that this is off by one from the traditional terminology of music theory, in which C–F is a "fourth". The line C–D–F can be thought of as the sequence of intervals C–D, D–F; similarly for any line.

Given two lines, we say they are related by a transposition in the diatonic if the sequences of diatonic lengths of their intervals are identical. This is easily seen to be an equivalence relation on the set of all lines. The equivalence classes will be called *genera*. For example, the genus containing C–D–F, which we denote $\langle$C–D–F$\rangle$, is

$$\{C–D–F, D–E–G, E–F–A, F–G–B, G–A–C, A–B–D, B–C–E\}.$$

Note that C–E–F is not in this genus; its sequence of lengths is 2, 1, which is not the same as 1, 2.

The *chromatic length* of an interval is the length of the interval measured in semitones; we shall use absolute value signs for chromatic length. Thus, $|C–F| = 5$. Given two lines in the same genus, we say they are related by a transposition in the chromatic if the sequences of chromatic lengths of their intervals are identical. This relation partitions each genus into subsets which we call *species*. For example, $\langle$C–D–F$\rangle$ partitions into the three species, $\{C–D–F, D–E–G, G–A–C, A–B–D\}$, $\{E–F–A, B–C–E\}$, and $\{F–G–B\}$. The lines in the first species all have intervals of chromatic lengths 2, 3; those of the second species, 1, 4; the lone line of the third species, 2, 4.

We see that the genus containing the *three*-note line C–D–F partitions into *three* species. This is not a coincidence, as the following theorem shows.

THEOREM 1. *Given any k, $1 \leqslant k \leqslant 7$, and any k-note line, the genus containing that line contains exactly k species.*

A brute-force proof can be carried out by examining all 13,699 lines, grouping them in their 1,957 genera, and counting the species within each genus. A more elegant proof, which leads to significant generalizations, is based on the construct known in music theory as the "circle of fifths." (See Fig. 2.)

The labels inside the upper semi-circle are diatonic lengths; the other labels are chromatic lengths. Measurement is clockwise.

Consider again the line C–D–F. The other lines in $\langle$C–D–F$\rangle$ are obtained by cycling around the upper semi-circle, since the diatonic distances on that semi-circle are constant. (See Fig. 3.)

The three species in this genus arise from the three possible locations of the short interval B–F (known in music theory as the "diminished fifth"); this interval can be included in the second interval of the three-note line (as in the first four diagrams), or in the first interval (next two

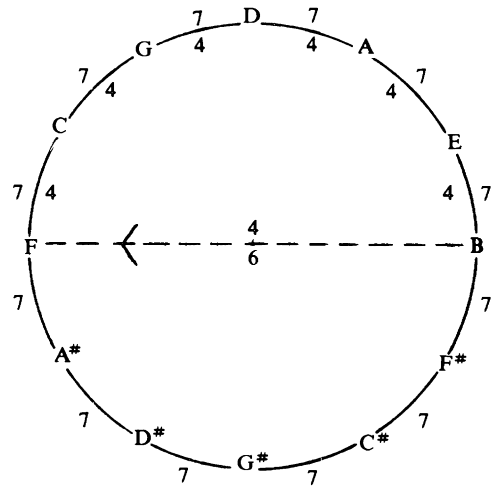


FIG. 2. The circle of fifths.

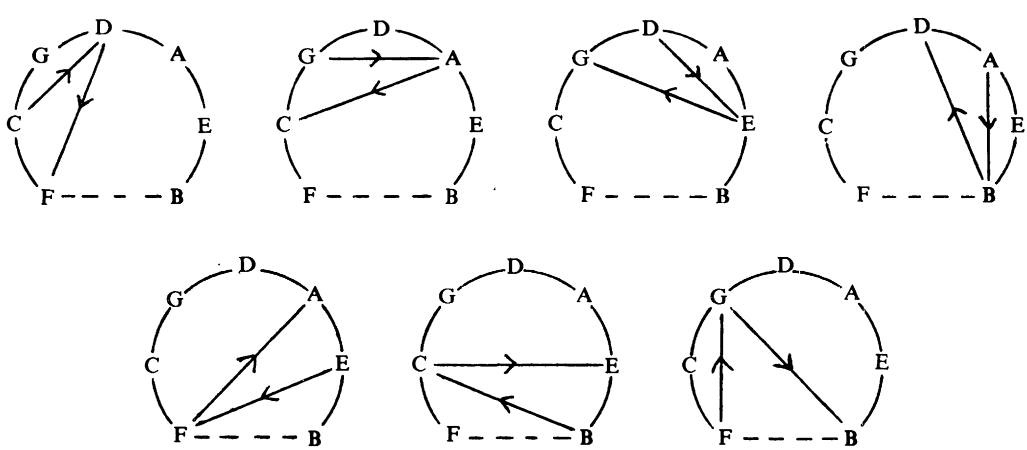


FIG. 3. The structure containing C–D–F.

diagrams), or in neither (last diagram). More generally, for any k, $1 \leqslant k \leqslant 7$, and any k-note line, the other lines in the genus are obtained by cycling through, and the species are distinguished by the position of B–F, for which there are k possibilities. This proves Theorem 1.

We can get somewhat more out of the circle of fifths. We have seen that the species in $\langle$C–D–F$\rangle$ contain 1, 2, and 4 lines. Now the distances from C to D, D to F, and F to C, when measured in fifths, are 2, 4, and 1, and the reason for the equality of these two sets of numbers is clear from Fig. 3. We state the general principle, which we call SM ("Structure yields Multiplicity"), in the following corollary to Theorem 1.

COROLLARY 1. *The numbers of lines in the species contained in the genus of a given line are given by the distances between each note in the line and the next note in the line, in the sense of clockwise travel around the circle of fifths. We include the distance from the last note to the first, and we measure distances in fifths.*

We leave it to the reader to define genus and species for chords, and to prove that given any k, $1 \leqslant k < 7$, and any k-note chord, the genus containing the chord contains exactly k species.

Myhill's property and its consequences. We now generalize. Instead of the usual chromatic set of twelve notes, we consider an abstract (but finite) chromatic set of c notes. By a *scale* we mean an ordered pair consisting of such a chromatic set together with a distinguished subset called the diatonic set. We let d be the cardinality of the diatonic set, and we label its elements $D_0, D_1, \ldots, D_{d-1}$. It is clear how we define line, interval, diatonic and chromatic lengths, genus, and species. We will make use later of the following simple property of lengths.

LEMMA 1. *Given k, $0 < k < d$, the sum of the chromatic lengths of the intervals of diatonic length k is ck.*

Proof. The number of semitones in the chromatic scale is, by definition, c, and each of these semitones is contained in precisely k of the intervals of diatonic length k.

A scale is said to have property CV ("Cardinality equals Variety") if for every k with $1 \leqslant k \leqslant d$, and for any k-note line, the genus containing that line contains exactly k species. This is the property of the usual scale asserted by Theorem 1. We wish to determine conditions under which a scale has property CV.

We define the *spectrum* of an interval I to be the set of all chromatic lengths of intervals in $\langle I \rangle$. If a scale has CV, then in particular every interval has a two-element spectrum. Our colleague John Myhill, to whom we are indebted for bringing us together to work on the problems of this paper, conjectured that the converse is true. We shall say that a scale has MP (Myhill's Property) if every interval has a two-element spectrum. We shall prove that MP implies CV by constructing, for any scale with MP, a "generalized circle of fifths."

Consider the spectrum of the interval $D_0 - D_1$. If this is a set of two consecutive integers, we say the scale is *rounded*. The usual scale is an example of a rounded scale. Given any scale in which the spectrum of $D_0 - D_1$ is a two-element set (in particular, any scale with MP), there is at least one corresponding rounded scale obtained by deleting non-diatonic notes, keeping equal chromatic lengths equal, as illustrated in Fig. 4. The deletion process preserves genera, since the diatonic set is unaffécted. It preserves species, since species membership rests on equality of chromatic lengths. Thus, a rounded scale so obtained has MP (or CV) if and only if the original scale does. For the remainder of this section, we assume all scales are rounded.

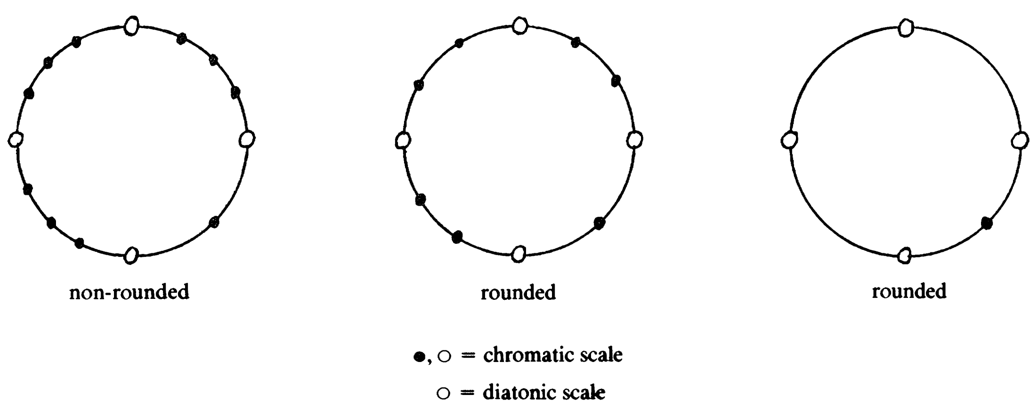


FIG. 4. Rounding a scale.

We say that a scale has CP (Consecutivity Property) if each interval has a spectrum consisting of consecutive integers.

LEMMA 2. *Every rounded scale has* CP.

Proof. Choose k, $0 < k < d$, and consider the interval $D_0 - D_k$. If the spectrum of this interval contains only one integer, the lemma is trivially true. Otherwise, there exist i and j with

$j - i = k$ and $|D_i - D_j| \neq |D_{i+1} - D_{j+1}|$, where the subscripts are to be reduced (mod d), if necessary, to bring them into the range $\{0, 1, \dots, d-1\}$. But

$$|D_{i+1} - D_{j+1}| - |D_i - D_j| = |D_j - D_{j+1}| - |D_i - D_{i+1}| = \pm 1,$$

since the scale is rounded. Thus no two consecutive terms in the sequence $|D_0 - D_k|$, $|D_1 - D_{k+1}|, \dots, |D_{d-1} - D_{k-1}|$ differ by more than one, and the elements of the spectrum are consecutive.

LEMMA 3. *If a scale with parameters c and d has* MP, *then* $(c, d) = 1$.

Proof. Suppose to the contrary a scale has MP and $(c, d) = r > 1$. Consider the genus $\langle D_0 - D_{d/r}\rangle$. There are d intervals in this genus, of total chromatic length $(d/r)c$ (by Lemma 1). Thus we have d integers summing to $(c/r)d$, and c/r is an integer. These d integers cannot all be c/r—if they were, the spectrum of $D_0 - D_{d/r}$ would have only the one element c/r, violating MP. Thus at least one of the integers exceeds c/r, and at least one falls short. To satisfy CP then, c/r must be in the spectrum, but then the spectrum has at least three elements, violating MP.

The following number-theoretical lemma is crucial to the construction of the generalized circle of fifths.

LEMMA 4. *Let* $(c, d) = 1$; *then there exists an integer* c', $0 \leqslant c' < d$, *such that* $cc' \equiv -1$ (mod d).

We omit the proof; a more general theorem is proved in the early chapters of nearly every introductory number theory textbook.

LEMMA 5. *Let a scale have* MP. *Let c' be as in the preceding lemma. Then with one exception the intervals of diatonic length c' all have chromatic length $d' = (cc' + 1)/d$; the exception has chromatic length $d' - 1$.*

Note that, for the usual scale, $c = 12$, $d = 7$, $c' = 4$, $d' = 7$, and the exceptional interval is the diminished fifth, B–F, with chromatic length $d' - 1 = 6$.

Proof. By Lemma 1, the sum of the chromatic lengths of the d intervals of diatonic length c' is cc'. By definition, $cc' = dd' - 1$, so we have d integers summing to $dd' - 1$. By MP there are exactly two distinct integers among these d integers, and by Lemma 2 they are consecutive; hence, $d - 1$ of these integers are d', and the other is $d' - 1$.

We can now label the diatonic set in such a way that

$$|D_0 - D_{c'}| = |D_{c'} - D_{2c'}| = \cdots = |D_{(d-2)c'} - D_{(d-1)c'}| = d', \quad |D_{(d-1)c'} - D_{dc'}| = d' - 1,$$

the subscripts being read (mod d). Thus we have constructed a generalized circle of fifths (Fig. 5).

THEOREM 2. MP *implies* CV.

Proof. The argument from the circle of fifths given for Theorem 1 above goes over to the generalized circle of fifths.

Property SM, of Corollary 1, is also relevant to the more abstract setting of this section. With the understanding that "fifths" is interpreted as "generalized fifths," we have

COROLLARY 2. MP *implies* SM.

Existence and uniqueness of scales with CV and given parameters. We now turn to the construction of scales with CV. Given parameters c, d with $(c, d) = 1$ we show that there exists a scale with CV with those parameters; moreover, it is essentially unique. We consider unicity first.

THEOREM 3. *Let S and S^* be scales with* CV *and with parameters c and d. Let their chromatic*

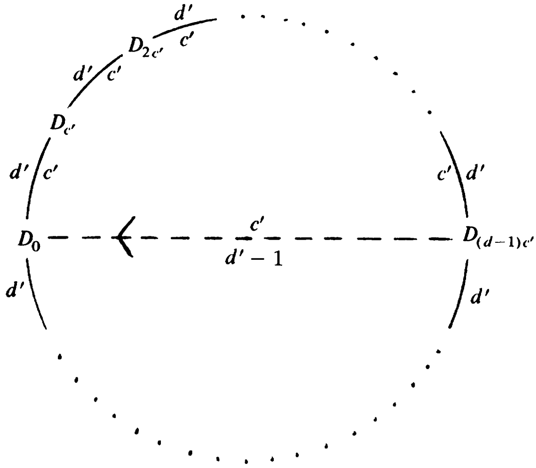


FIG. 5. Generalized circle of fifths.

sets be

$$C = \{C_0, C_1, \ldots, C_{c-1}\} \quad \textit{and} \quad C^* = \{C_0^*, C_1^*, \ldots, C_{c-1}^*\},$$

respectively, and let their diatonic sets be D *and* D^*, *respectively. Then there exists an integer* j *such that* C_i *is in* D *if and only if* C_{i+j}^* *is in* D^*; *here, the subscripts are to be interpreted* (modulo c).

Proof. Since the scales share parameters c and d, they also share c' and d'. By the construction of the generalized circle of fifths, there are notes C_k in C and C_h^* in C^* such that

$$D = \{C_k, C_{k+d'}, \ldots, C_{k+(d-1)d'}\} \quad \text{and} \quad D^* = \{C_h^*, C_{h+d'}^*, \ldots, C_{h+(d-1)d'}^*\}.$$

Now simply let $j = h - k$.

Concerning the existence of scales with CV and given parameters, we first show how to construct the usual scale. Here $c = 12$ and $d = 7$. Write down the multiples of $\frac{12}{7}$: $0, 1\frac{5}{7}, 3\frac{3}{7}, 5\frac{1}{7}, 6\frac{6}{7}, 8\frac{4}{7}, 10\frac{2}{7}, 12, \ldots$. Then erase the fractions, leaving $0, 1, 3, 5, 6, 8, 10, 12, \ldots$. Interpret this sequence as the positions of the white keys, and the omitted integers $2, 4, 7, 9, 11, \ldots$ as the positions of the black keys, and you obtain a scale with CV; if you identify position 0 with the note B, you recover the usual (C-major) scale:

0	1	2	3	4	5	6	7	8	9	10	11	12	...
B	C	C♯	D	D♯	E	F	F♯	G	G♯	A	A♯	B	...

FIG. 6

We will prove that this procedure works quite generally. First we need to quote another well-known lemma from elementary number theory.

LEMMA 6. *If* r, s, t *are integers,* r *divides* st, *and* $(r, s) = 1$, *then* r *divides* t.

THEOREM 4. *Given* c *and* d *with* $(c, d) = 1$, *let* $a_k = \left[\frac{kc}{d}\right]$, $k = 0, \pm 1, \pm 2, \ldots$. *Then the integers* a_k *are the positions of the notes of the diatonic set in a scale with* CV *with parameters* c *and* d.

Proof. It suffices to show that the scale so constructed has MP, that is, for every j, $1 \leqslant j < d$, the set $\{a_{k+j} - a_k : k = 0, \pm 1, \pm 2, \ldots\}$ has cardinality two. Since, for all x, $x - 1 < [x] \leqslant x$,

and since $a_{k+j} - a_k = \left[\frac{(k+j)c}{d}\right] - \left[\frac{kc}{d}\right]$, we have

$$\frac{(k+j)c}{d} - 1 - \frac{kc}{d} < a_{k+j} - a_k < \frac{(k+j)c}{d} - \frac{kc}{d} + 1.$$

Thus for fixed j there is an open interval (in the mathematical sense) of length 2 containing the spectrum of the interval (in the musical sense) of diatonic length j. Such a mathematical interval contains at most two integers, so the spectrum is at most a two-element set. Now suppose the spectrum is a one-element set. Then there is an integer, call it e, such that, for all k, $a_{k+j} - a_k = e$. Then

$$de = \sum_{n=1}^{d} \left(a_{k+nj} - a_{k+(n-1)j}\right) = a_{k+dj} - a_k = \left[\frac{(k+dj)c}{d}\right] - \left[\frac{kc}{d}\right] = cj,$$

so d divides cj. By Lemma 6, d divides j. But $1 \leqslant j < d$ by assumption, yielding a contradiction. Thus, the spectrum is a two-element set, so the scale has MP.

Concluding remarks. While the connection between mathematics and music goes back almost as far as the beginnings of the two disciplines themselves, the features of musical structure studied in this paper have not heretofore been subjected to rigorous analysis. (The reader will, however, find the contents of [**1**], [**2**], [**3**], and [**4**] to be relevant to our discussions.) We hope that our results will stimulate further research on mathematics and music theory.

References

1. Richmond Browne, Tonal implications of the diatonic set, In Theory Only, 5/6–7 (1981) 3–21.
2. John Clough, Aspects of diatonic sets, J. Music Theory, 23/1 (1979) 45–61.
3. ______, Diatonic interval sets and transformational structures, Perspectives of New Music, 18/2 (1980) 461–482.
4. Carlton Gamer, Some combinational resources of equal-tempered systems, J. Music Theory, 11/1 (1967) 32–59.
5. David L. Reiner, Enumeration in music theory, this MONTHLY, 92 (1985) 51–54.

MORE TREES AND POWER SUMS

JAMES E. SIMPSON
Department of Mathematics, University of Kentucky, Lexington, KY 40506

1. Introduction. A recent article in the MONTHLY by A. H. Stone [**8**] gives an engaging presentation of the formula

$$n^{n-3} = \frac{1}{2}\sum_{r=1}^{n-1}\binom{n-2}{r-1} r^{r-2}(n-r)^{n-r-2} \quad \text{for } n \geq 2 \tag{1}$$

as a plausible generalization of two well-known equations, $5^2 = 4^2 + 3^2$ and $6^3 = 5^3 + 4^3 + 3^3$. The fundamental combinatorial technique Stone uses to prove (1) is that of counting a particular set in two ways, equating the results. Here we use this same technique of double counting to give

James E. Simpson: I am a native of Chicago, with a Ph. D. from Yale University (1961). My research area was originally in analysis, but after a year of close association with Professor M. Behzad in 1972 I have floated gradually into combinatorics and graph theory. My major project this year is to finish an introductory text in discrete mathematics.

a somewhat simpler proof of (1). We also take a brief excursion through the Matrix-Tree Theorem and the Principle of Inclusion-Exclusion to show how to develop a whole family of similar equalities. Most of these are special cases of so-called Abel identities.

2. Graphical background. We require only simple graphs for this discussion. So $G = (V, E)$ is a graph with vertex set $V = V(G)$ and edge set $E = E(G)$ with no loops or multiple edges. Also, all of our graphs are connected, so that between every pair of vertices there is at least one path. A connected graph with no simple closed paths (cycles) is a tree. For any graph G, a graph T is said to be a spanning tree of G if T is a tree for which $V(T) = V(G)$ and $E(T) \subseteq E(G)$. It is well known that any tree on n vertices has $n - 1$ edges. A given graph G may have many spanning trees, and we denote by $\tau(G)$ the number of these spanning trees. A famous theorem of Cayley asserts that $\tau(K_n) = n^{n-2}$, where K_n is the complete graph on n vertices. Since K_n has an edge for every pair of vertices, it is obvious that

$$|E(K_n)| = \binom{n}{2} = \frac{1}{2}n(n-1).$$

Proofs of Cayley's Theorem, based on Prüfer codes, may be found in [**8**] and in [**1**, p. 35]. The latter reference is a good one for general information on graphs.

3. Proving (1). Fix an edge e_j of K_n and let B_j be the set of spanning trees of K_n that use e_j. Observe that $|B_j|$ is the same for every e_j, $1 \le j \le \binom{n}{2}$. For if e_i and e_j are any two edges of K_n, and σ is any permutation of the vertices of K_n that carries the endpoints of e_i to the endpoints of e_j, then σ immediately provides a matching between B_i and B_j. Formula (1) is obtained by counting $|B_j|$ in two ways. One way, used by Stone in [**8**], relies on the observation that for a fixed edge, any tree T in B_j may be regarded as the union of e_j with two subtrees, one with each end of e_j in its vertex set. So if $e_j = (u, v)$, then T is the union of a tree T_1 with $u \in V(T_1)$, a tree T_2 with $v \in V(T_2)$, and the edge e_j itself. We may construct a typical tree in B_j by choosing an integer r, $1 \le r \le n-1$, a subset R of $V(K_n)$ of size r that contains u, a tree T_1 using R as its vertex set, and a tree T_2 using the remaining $n - r$ vertices of K_n as *its* vertex set. By Cayley's Theorem, T_1 may be chosen in r^{r-2} ways and T_2 in $(n-r)^{n-r-2}$ ways. These observations taken together yield

$$|B_j| = \sum_{r=1}^{n-1} \binom{n-2}{r-1} r^{r-2}(n-r)^{n-r-2}.$$

There is a second way to count $|B_j|$ that is much simpler than the one in [**8**]. Again it involves counting a set in two ways. Let N be the number of pairs (e, T) for which T is a spanning tree of K_n and e is an edge in $E(T)$. If we count the trees first, then the edges in each T, we obtain $N = n^{n-2}(n-1)$. If we count the edges first, then the trees using each edge, we obtain $N = \binom{n}{2}|B_j|$. Setting these values of N equal to each other and solving for $|B_j|$, we have the desired second value, $|B_j| = 2n^{n-3}$. Equating the two values of $|B_j|$ yields (1).

An equivalent form of (1) is an old identity of Abel. Rewriting (1) with $n + 2$ in place of n and changing the summation index to $k = r - 1$, we have

$$(n+2)^{n-1} = \frac{1}{2}\sum_{k=0}^{n}\binom{n}{k}(k+1)^{k-1}(n-k+1)^{n-k-1}. \tag{2}$$

Riordan [**6**] gives a general Abel formula:

$$A_n(x, y; p, q) = \sum_{k=0}^{n}\binom{n}{k}(x+k)^{k+p}(y+n-k)^{n-k+q},$$

and [**6**, p. 23] the identity

$$A_n(x, y; -1, -1) = (x^{-1} + y^{-1})(x + y + n)^{n-1},$$

which reduces to (2) for $x = y = 1$.

4. Variations. Let A_j be the set of spanning trees of K_n that avoid e_j, so that for each j, (A_j, B_j) is a partition of the set of all spanning trees of K_n. From the work just done,

$$|A_j| = \tau(K_n) - |B_j| = n^{n-2} - 2n^{n-3} = n^{n-3}(n-2).$$

As Stone remarks, this formula may also be obtained directly from the Matrix-Tree Theorem, which we restate below. We will use it to count the number of spanning trees that avoid certain selected sets of edges of K_n. For an arbitrary graph G, suppose the vertices and edges of G have been labelled in some convenient order, so that

$$V(G) = \{v_1, v_2, \dots, v_q\} \quad \text{and} \quad E(G) = \{e_1, e_2, \dots, e_s\}.$$

Next assign an arbitrary direction to each edge. Then each $e_j = (u, v)$ is an ordered pair with u as the tail of e_j and v as the head. Define M to be the vertex-edge incidence matrix whose (i, j)-entry $m_{i,j}$ is 1 if v_i is the head of e_j, -1 if v_i is the tail of e_j, and 0 otherwise. Let M^t be the transpose of M. The (i, j)-entry of MM^t is the dot product of the ith and jth rows of M. Let Q be obtained from MM^t by deleting the ith row and column, for some choice of i. (In fact it is immaterial what i is deleted. We use $i = q$.) Then Q is a $(q-1)$ by $(q-1)$ matrix. From the way M is defined, if $i \neq j$ then $q_{i,j}$, the (i, j)-entry of Q, is -1 if v_i and v_j are the endpoints of some edge of G, otherwise $q_{i,j} = 0$. For $i = j$, $q_{i,i}$ is the degree of v_i, the number of edges of which v_i is an endpoint.

MATRIX-TREE THEOREM. *For any graph G, $\tau(G) = \det(Q)$.*

Proofs may be found in [**1**, p. 219] and in [**4**, p. 221].

When $G = K_n$, Q has all off-diagonal entries -1, and all diagonal entries $n - 1$:

$$Q = \begin{vmatrix} n-1 & & & \\ & & & -1 \\ & n-1 & & \\ -1 & & \ddots & \\ & & & n-1 \end{vmatrix}.$$

Evaluating the determinant of this Q gives another way to prove Cayley's Theorem. (See [**4**, p. 222].) We show some details as preparation for evaluating $\det(Q)$ for a different G a little further on. We use elementary row operations to calculate $\det(Q)$ as follows. First add every row except the first to row one. Since each column has $n - 2$ entries of -1, the result is

$$\det(Q) = \det\begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ -1 & n-1 & -1 & \dots & -1 \\ -1 & -1 & n-1 & \dots & -1 \\ & & & \ddots & \\ -1 & -1 & \dots & & n-1 \end{bmatrix}.$$

Next add row one to each of the other rows to obtain

$$\det(Q) = \det\begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 0 & n & 0 & \dots & 0 \\ 0 & 0 & n & \dots & 0 \\ \dots & & & & \\ 0 & 0 & 0 & \dots & n \end{bmatrix} = n^{n-2},$$

since the determinant is now just the product of the diagonal entries.

Our next application is to graphs obtained from K_n by deleting certain edges. A subset, S, of $E(K_n)$ is a *partial matching* if the edges in S share no endpoints. Let G_S be the graph obtained from K_n by deleting the edges of S. Clearly, if $|S| = t$, then $t \leq n/2$.

THEOREM (Weinberg [9]). *For any partial matching, S, $\tau(G_S) = n^{n-2-t}(n-2)^t$.*

Proof. First label the vertices and edges of K_n so that $S = \{e_1, e_2, \ldots, e_t\}$ and the endpoints of e_j are v_{2j-1} and v_{2j} for $j = 1, 2, \ldots, t$. In the incidence matrix M for G_S, for rows $2j$ and $2j-1$, there are no columns in which both of these rows have non-zero entries simultaneously, for $j = 1, 2, \ldots, t$. To compute Q we delete the last row and column of MM^t. Then

$$
\begin{array}{rc}
 & \qquad\qquad\qquad\qquad (\text{col } 2t) \\
\begin{array}{r} Q = \\ \\ (\text{row } 2t) \end{array} &
\left|\begin{array}{cc|cc|ccc|ccc}
n-2 & 0 & -1 & -1 & \cdots & & & & & -1\\
0 & n-2 & -1 & -1 & \cdots & & & & & -1\\ \hline
-1 & -1 & n-2 & 0 & & -1 & & & -1 & \\
-1 & -1 & 0 & n-2 & & & & & & \\ \hline
 & & & & \ddots & & & & & \\
 & -1 & & -1 & & n-2 & 0 & & -1 & \\
 & & & & & 0 & n-2 & & & \\ \hline
 & & & & & & & n-1 & & \\
 & & & & & & & & & -1 \\
 & & & -1 & & & & & n-1 & \\
 & & & & & & & & & \ddots \\
 & & & & & & & -1 & & n-1
\end{array}\right|
\end{array}
$$

We use the same two transformations as before to evaluate $\det(Q)$. Add all rows to the first row:

$$
\begin{array}{rc}
 & \qquad\qquad\qquad\qquad (\text{col } 2t) \\
\begin{array}{r} \det(Q) = \det \\ \\ (\text{row } 2t) \end{array} &
\left|\begin{array}{cc|cc|ccc|ccc}
1 & 1 & 1 & 1 & \cdots & & & & & \\
0 & n-2 & -1 & -1 & \cdots & & & & & -1\\ \hline
 & -1 & n-2 & 0 & & -1 & & & -1 & \\
 & & 0 & n-2 & & & & & & \\ \hline
 & & & & \ddots & & & & & \\
 & -1 & & -1 & & n-2 & 0 & & -1 & \\
 & & & & & 0 & n-2 & & & \\ \hline
 & & & & & & & n-1 & & -1 \\
 & & & -1 & & & & & \ddots & \\
 & & & & & & & -1 & & n-1
\end{array}\right|
\end{array}
$$

Adding row one to every other row gives

$$
\begin{array}{rc}
 & \qquad\qquad\qquad\qquad (\text{col } 2t) \\
\begin{array}{r} \det(Q) = \det \\ \\ (\text{row } 2t) \end{array} &
\left|\begin{array}{cc|cc|ccc|ccc}
1 & 1 & 1 & 1 & 1 & \cdots & & & & 1\\
1 & n-1 & 0 & 0 & 0 & \cdots & & & & 0\\ \hline
0 & 0 & n-1 & 1 & & 0 & & & & \\
0 & 0 & 1 & n-1 & & & & & 0 & \\ \hline
 & & & & \ddots & & & & & \\
 & 0 & & & & n-1 & 1 & & & \\
 & & & & & 1 & n-1 & & & \\ \hline
 & & & & & & & n & & 0 \\
 & & & 0 & & & & & \ddots & \\
 & & & & & & & 0 & & n
\end{array}\right|
\end{array}
$$

This determinant is the product of t 2 by 2 determinants along the main diagonal, multiplied by n^{n-1-2t}. The first 2 by 2 determinant has value $n-2$, and the remaining $t-1$ of them have value $(n-1)^2-1$. So

$$\det(Q) = (n-2)(n^2-2n)^{t-1}n^{n-1-2t} = n^{n-2-t}(n-2)^t.$$

This theorem appears with an indication of the proof above in [**5**, p. 44]. This reference has many more results about trees, as well as a general form of the Abel Identity given above. It also contains a theorem for counting those trees that contain certain special subsets of $E(K_n)$, in [**5**, Theorem 6.1, p. 52]. We use inclusion-exclusion to derive a special case.

THEOREM. *As before, let S be a partial matching in K_n, with $|S| = t$. Then the number of spanning trees of K_n, each of which uses all the edges of S, is $2^t n^{n-2-t}$.*

Proof. We have already seen the case $t=1$ in the course of proving (1), in showing $|B_j| = 2n^{n-3}$. For arbitrary t, if J is any subset of $\{1,2,\dots,t\}$, with $|J| = r$, we let

$$A_J = \bigcap_{i\in J} A_i.$$

By the theorem just proved,

$$|A_J| = n^{n-2-r}(n-2)^r.$$

A general form of the inclusion-exclusion principle, valid whenever A_i and B_i are complementary subsets of the same finite set for each i, asserts that

$$\left|\bigcap_{i=1}^{t} B_i\right| = \textstyle\sum_J (-1)^{|J|}|A_J|,$$

where the summation extends over all subsets of $\{1,2,\dots,t\}$. Since for each r there are $\binom{t}{r}$ subsets J of size r, we have

$$\left|\bigcap_{i=1}^{t} B_i\right| = \sum_{r=0}^{t}(-1)^r\binom{t}{r}n^{n-2-r}(n-2)^r.$$

Finally, we factor n^{n-2-t} to the front and use the binomial theorem to obtain

$$\begin{aligned}\left|\bigcap_{i=1}^{t} B_i\right| &= n^{n-2-t}\sum_{r=0}^{t}(-1)^r\binom{t}{r}n^{t-r}(n-2)^r\\ &= n^{n-2-t}(n-(n-2))^t\\ &= 2^t n^{n-2-t}.\end{aligned}$$

We are now in a position to use this theorem to produce the analogues of (1) promised at the beginning. For $t=2$ and $n \geq 4$, we have $|B_1 \cap B_2| = 4n^{n-4}$. Let us count this set in another way, by decomposing each tree into a union of subtrees, much as in Section 3 above. Suppose T is a tree that uses both $e_1 = (u, v)$ and $e_2 = (x, y)$. Removing e_1 from T gives two subtrees, T_1 using u, say, and T_2 using v. One of these subtrees includes e_2. Accordingly, we partition $B_1 \cap B_2$ into the union of two disjoint subsets C_1 and C_2, where C_1 contains every $T = T_1 \cup T_2 \cup \{e_1\}$ for which e_2 is in T_1 and C_2 has those for which e_2 is in T_2. By the symmetry of K_n, it is clear that $|C_1| = |C_2|$. Let us count C_2. An arbitrary tree in C_2 may be obtained by choosing any set of r vertices that includes u and omits v, x and y to be the vertex set of T_1. This can be done in $\binom{n-4}{r-1}$ ways. The remaining $n-r$ vertices are used for T_2. There are r^{r-2} possible choices for T_1. The possible number of T_2's *that use* e_2 is $2(n-r)^{n-r-3}$. Therefore,

$$|B_1 \cap B_2| = 2|C_2| = 2\sum_{r=1}^{n-3}\binom{n-4}{r-1}r^{r-2}2(n-r)^{n-r-3}.$$

Equating these two evaluations of $|B_1 \cap B_2|$ we have

$$(3) \qquad n^{n-4} = \sum_{r=1}^{n-3} \binom{n-4}{r-1} r^{r-2}(n-r)^{n-r-3} \quad \text{for } n \geq 4.$$

As with (2) we may transform (3) by substituting $n + 4$ for n and changing the index. Then the appropriate Abel identity [6, p. 23] is

$$A_n(x, y; -1, 0) = x^{-1}(x + y + n)^n,$$

which becomes (3) when we set $x = 1$ and $y = 3$.

We indulge ourselves in one further case, based on $t = 3$, $n \geq 6$. Then

$$|B_1 \cap B_2 \cap B_3| = 8n^{n-5}.$$

Let $e_1 = (u, v)$, $e_2 = (x, y)$ and $e_3 = (w, z)$. Now each T in $B_1 \cap B_2 \cap B_3$ may be written $T_1 \cup T_2 \cup \{e_1\}$ with four possibilities:

C_1: Both e_2 and e_3 are in T_1.
C_2: Both are in T_2.
C_3: e_2 is in T_1 and e_3 is in T_2.
C_4: e_3 is in T_1 and e_2 is in T_2.

Again by symmetry, $|C_1| = |C_2|$ and $|C_3| = |C_4|$. By techniques just like those above,

$$|C_2| = \sum_{r=1}^{n-5} \binom{n-6}{r-1} r^{r-2} 4(n-r)^{n-r-4},$$

and

$$|C_3| = \sum_{r=3}^{n-3} \binom{n-6}{r-3} 2r^{r-3} 2(n-r)^{n-r-3}.$$

Using $|B_1 \cap B_2 \cap B_3| = 2(|C_2| + |C_3|)$ and equating the results yields

$$(4) \qquad n^{n-5} = \sum_{r=1}^{n-5} \binom{n-6}{r-1} \left[r^{r-2}(n-r)^{n-r-4} + (r+2)^{r-1}(n-r-2)^{n-r-5} \right]$$

for $n \geq 6$.

As before, substituting $n + 6$ for n and changing the index, we have

$$(5) \quad (n+6)^{n+1} = \sum_{k=0}^{n} \binom{n}{k} \left[(k+1)^{k-1}(n-k+5)^{n-k+1} + (k+3)^k (n-k+3)^{n-k} \right]$$
$$= A_n(1, 5; -1, 1) + A_n(3, 3; 0, 0).$$

While this does not appear in [6], the identities

$$A_n(x, y; -1, 1) = x^{-1} \sum_{k=0}^{n} \binom{n}{k} (x + y + n)^{n-k} k!(y + k)$$

and

$$A_n(u, v; 0, 0) = \sum \binom{n}{k} (u + v + n)^{n-k} k!$$

do appear. From these, whenever $x + y = u + v$, we have

$$A_n(x, y; -1, 1) + A_n(u, v; 0, 0) = x^{-1} \sum_{k=0}^{n} \binom{n}{k} (x + y + n)^{n-k} k!(x + y + k).$$

Riordan [6, p. 21] also has

$$(x+y+n)^{n+1} = \sum_{k=0}^{n} \binom{n}{k} k!(x+y+k)(x+y+n)^{n-k},$$

so that for $x + y = u + v$:

$$A_n(x, y; -1,1) + A_n(u, v; 0,0) = x^{-1}(x+y+n)^{n+1}.$$

This reduces to (5) with $x = 1$, $y = 5$ and $u = v = 3$.

The same kind of argument, details of which we leave to the energetic reader, using $t = 4$ and $n \geq 8$, produces

$$(6) \qquad n^{n-6} = \sum_{r=1}^{n-7} \binom{n-8}{r-1} \left[r^{r-2}(n-r)^{n-r-5} + 3(r+2)^{r-1}(n-r-2)^{n-r-6} \right].$$

This may be rewritten as

$$(n+8)^{n+2} = A_n(1,7; -1,2) + 3A_n(3,5;0,1), \quad n \geq 0.$$

As the value of t increases, the right sides of these equalities become more complicated and less graceful. Only (1) is pleasant enough to provide for two terms on the right with non-fractional coefficients for the power terms. Nonetheless there is a general form that we now demonstrate.

THEOREM. *For n and $q \geq 0$,*

$$(7) \qquad (n+2q+2)^{n+q-1} = \frac{1}{2} \sum_{s=0}^{q} \binom{q}{s} A_n(2s+1, 2q-2s+1; s-1, q-s-1).$$

Proof. Let S, with $|S| = t = q + 1$, be a partial matching in K_N where $N = 2t + n$. As usual we count the spanning trees of K_N that use S. We already know the total,

$$2^t(N)^{N-2-t} = 2^t(n+2t)^{n+t-2}$$

Letting $S = \{e_1, \ldots, e_t\}$ and $e_1 = (u, v)$ we regard any spanning T as the union of e_1 with two subtrees, T_1 using u and T_2 using v. Any such tree may be constructed by first choosing the number s of elements of S to be in T_1, with $0 \leq s \leq q = t - 1$. Next we choose a subset of $S \setminus \{e_1\}$ of size s in $\binom{q}{s}$ ways. Counting u we now choose $k + 1$ additional vertices to be in T_1 in $\binom{n}{k}$ ways, $0 \leq k \leq n$. From the previous theorem the number of ways to build T_1 on these $2s + k + 1$ vertices using the chosen edges is $2^s(2s + k + 1)^{s+k-1}$. Finally, T_2 uses the remaining $q - s$ edges of $S \setminus \{e_1\}$ and the remaining $n - k + 1$ vertices, including v. Again by the previous theorem T_2 may be built in

$$2^{q-s}(n+2q-2s-k+1)^{n+q-s-k-1}$$

ways. Putting this all together we have

$$2^{q+1}(n+2q+2)^{n+q-1} = \sum_{s=0}^{q} \binom{q}{s} \sum_{k=0}^{n} 2^q(2s+k+1)^{s+k-1}(n+2q-2s-k+1)^{n-k+q-s-1}.$$

After division by 2^{q+1} the last terms of the summation may be translated into the appropriate A_n's to obtain (7).

4. Remarks. The author does not know if the equalities (6) and (7) are original. But even if they are known, the interplay of various combinatorial ideas in their development is worthwhile. There are also connections to major areas of interest in the field. One proof of the Matrix-Tree Theorem, for example, leads directly to the topic of matroids. The Abel identities are related to techniques of manipulation of formal series embodied in the so-called umbral calculus. Indeed, the search for combinatorial proofs of general identities has received some attention in recent years. The articles in [3] and [7] are good examples of this work. To any reader who has found the

development given in this paper intriguing, we especially recommend [2]. In that paper, Broder gives an analogous, more powerful development of Abel identities for all $p, q \geq 0$. Briefly, instead of using the Prüfer code for trees, a more complex coding of a special class of directed graphs is analyzed, to obtain the identities.

Finally, we wish to thank those colleagues who have commented on the connection of this work to Abel identities (e.g., see a letter from I. Gessel to this MONTHLY, 93 (1986) 323).

References

1. J. A. Bondy and U. S. R. Murty, Graph Theory with Applications, North Holland, New York, 1976.
2. A. Z. Broder, A general expression for Abelian identities, in: L. J. Cummings, ed., Combinatorics on Words, Progress and Perspectives, Academic Press, 1983, 229–245.
3. S. A. Joni, G.-C. Rota, and B. Sagan, From sets to functions: three elementary examples, Discrete Math., 37 (1981) 193–202.
4. L. Lovasz, Combinatorial Problems and Exercises, North Holland, 1979.
5. J. W. Moon, Counting Labelled Trees, Canad. Math. Monographs, No. 1, 1970.
6. J. Riordan, Combinatorial Identities, Krieger, 1979.
7. B. E. Sagan, A note on Abel polynomials and rooted labelled forests, Discrete Math., 44 (1983) 293–298.
8. A. H. Stone, Trees and power-sums, this MONTHLY, 92 (1985) 328–331.
9. L. Weinberg, Number of trees in a graph, Proc. IRE, 46 (1958) 1954–1955.

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

QUESTIONS AND CONJECTURES IN PARTITION THEORY

GEORGE E. ANDREWS
Department of Mathematics, The Pennsylvania State University, University Park, PA 16802

In order to present several unsolved problems in a way that connects them with what is already known and with the historical origins of the questions, we begin with an account of these background ideas. We also provide some key references.

The classical generating functions of partition theory provide two types of behavior in the growth of the coefficients of the power series. For example, in

$$(1) \qquad 1+\sum_{n=1}^{\infty}\frac{q^{n^2}}{(1-q)^2(1-q^2)^2\cdots(1-q^n)^2}=\prod_{n=1}^{\infty}(1-q^n)^{-1}=\sum_{n=0}^{\infty}p(n)q^n$$

(see [3; p. 21, eq. (2.2.9)]) the coefficients $p(n)$, the number of partitions of n, blow up, i.e., $p(n)\to\infty$ as $n\to\infty$. On the other hand, we find in Euler's famous Pentagonal Number Theorem (see [3; p. 11, eq. (1.3.1) and p. 19, eq. (2.2.6)]),

(2)
$$1+\sum_{n=1}^{\infty}\frac{(-1)^n q^{n(n+1)/2}}{(1-q)(1-q^2)\cdots(1-q^n)}=\prod_{n=1}^{\infty}(1-q^n)=1+\sum_{n=1}^{\infty}(-1)^n q^{n(3n-1)/2}(1+q^n)$$

that the power series coefficients are bounded; indeed they are all 0, ± 1.

We note in passing that the unboundedness of $p(n)$ and the truth of (2) (which tacitly implies the boundedness of the coefficients in (2)) are easily obtained by combinatorial methods (see [7] and [**3**; Chap. 1]).

Indeed a lengthy search of some of the literature (e.g. [**1**], [**2**], [**3**], [**6**], [**9**], [**10**]) on partitions reveals that most of the q-series considered either have coefficients which tend to infinity in absolute value or are bounded.

In studying Ramanujan's "Lost" Notebook [**4**], [**5**], I have come across some strikingly simple-looking q-series where the behavior appears to be quite different from that described above. This observation has raised a number of problems that might be accessible by combinatorial means.

Consider

$$\sigma(q) = 1 + \sum_{n=1}^{\infty} \frac{q^{n(n+1)/2}}{(1+q)(1+q^2)\cdots(1+q^n)} = \sum_{n=0}^{\infty} S(n) q^n$$

$$= 1 + q - q^2 + 2q^3 + \cdots + 4q^{45} + \cdots + 6q^{1609} + \cdots + 8q^{3288} + \cdots . \tag{3}$$

Actually it appears that a majority of the $S(n)$ are equal to 0. For various j, Table 1 shows the fraction of $n \leqslant j$ such that $S(n) = 0$.

TABLE 1

j	100	200	300	400	500	600	700	800	900	1000
fraction	0.44	0.475	0.51	0.5175	0.534	0.547	0.554	0.556	0.557	0.562

Let us make two conjectures.

CONJECTURE 1. $\limsup |S(n)| = +\infty$.

CONJECTURE 2. $S(n) = 0$ *for infinitely many* n.

Now how can one attack such problems? Perhaps classical saddle point methods could be used here [**8**], [**3**; Chap. 6]; however (3) suggests that the $S(n)$ are not very big. Hence it may be that a combinatorial approach is required. How would this go:

The rank of partition is the largest part minus the number of parts. Thus the rank of $10 + 9 + 3 + 2 + 1$ is 5. It is an elementary exercise to show that $S(n)$ is the excess of the number of partitions of n into distinct parts with even rank over those with odd rank. Thus $S(8) = -2$ since, of the 6 partitions of 8 with distinct parts, two have even rank $(6 + 2, 5 + 2 + 1)$ while four have odd rank $(8, 7 + 1, 5 + 3, 4 + 3 + 1)$.

Can one find a "near bijection" between these two types of partitions of n with distinct parts? This is the way F. Franklin proved Euler's Pentagonal Number Theorem [**3**; Chap. 1, pp. 10–11]. However Franklin's "near bijection" almost always is a bijection. Things will obviously be more complicated here.

There are other seemingly innocent q-series where it seems likely from numerical evidence that the coefficients are getting large; however often the behavior of the terms is quite variable. Let us begin with another example from Ramanujan's "Lost" Notebook [**4**; p. 57, eq. (1.10)]:

$$v_1(q) = 1 + \sum_{n=1}^{\infty} \frac{q^{n(n+1)/2}}{(1+q^2)(1+q^4)\cdots(1+q^{2n})} = \sum_{n=0}^{\infty} V_1(n) q^n. \tag{4}$$

CONJECTURE 3. $|V_1(n)| \to \infty$ *as* $n \to \infty$.

However the growth of $|V_1(n)|$ is not very smooth: some sample values of $V_1(n)$ are shown in Table 2.

TABLE 2

n	197	198	199	292	293	294	411	412	413
$V_1(n)$	33	0	-29	-367	4	375	473	4	-497
n	544	545	546	703	704	705	876	877	878
$V_1(n)$	6195	-18	-6309	-8550	-224	8716	-114717	54	116617

Again it would appear that smooth asymptotic results accessible by standard analytic means are questionable here. Perhaps the combinatorial interpretation of $V_1(n)$ described below could assist. There does however appear to be great sign regularity in $V_1(n)$:

CONJECTURE 4. *For almost all n, $V_1(n)$, $V_1(n+1)$, $V_1(n+2)$ and $V_1(n+3)$ are two positive and two negative numbers.*

CONJECTURE 5. *For $n \geqslant 5$ there is an infinite sequence $N_5 = 293$, $N_6 = 410$, $N_7 = 545$, $N_8 = 702, \ldots, N_n \geqslant 10n^2, \ldots$ such that $V_1(N_n)$, $V_1(N_n+1)$, $V_1(N_n+2)$ all have the same sign.*

CONJECTURE 6. *With reference to Conjecture 5, the numbers $|V_1(N_n)|$, $|V_1(N_n+1)|$, $|V_1(N_n+2)|$ contain a local minimum of the sequence $|V_1(j)|$.*

For example $N_8 = 702$, $V_1(N_8) = -273$, $V_1(N_8+1) = -8550$, $V_1(N_8+2) = -224$. Furthermore it is true that $|V_1(n)| \geqslant 224$ for $557 \leqslant n \leqslant 876$ with equality only at $n = 704$.

Examination of the values of $V_1(n)$ for $n \leqslant 2500$ shows the validity of these conjectures in this range.

Now $V_1(n)$ is closely related to partitions with alternating parity. In [**4**], we centered our attention on results from Ramanujan's "Lost" Notebook that concerned $OE(n)$, the number of partitions of n in which the parity of the parts alternates from smallest to largest part with the smallest part odd. Thus $OE(9) = 3$, since the admissible partitions are $9, 1+8, 3+6$. For brevity we call such partitions *odd-even* partitions.

The rank of each odd-even partition is clearly even; hence half the rank of such partitions is always an integer. Elementary arguments show that $V_1(n)$ is the number of odd-even partitions of n with rank $\equiv 0 \pmod 4$ minus the number with rank $\equiv 2 \pmod 4$. Are there any combinatorial mappings between these two classes of partitions that explain or partially explain Conjectures 4–6?

In addition to $v_1(q)$, the following series (taken from equations (1.8)–(1.11) in [**4**]) also appear to satisfy Conjecture 3:

$$v_2(q) = \sum_{n=1}^{\infty} \frac{q^{2n^2-n}}{(1+q)(1+q^3)\cdots(1+q^{2n-1})} = \sum_{n=0}^{\infty} V_2(q)q^n, \tag{5}$$

$$v_3(q) = 1 + \sum_{n=1}^{\infty} \frac{(-1)^n q^{n(n+1)/2}}{(1+q)^2(1+q^2)^2\cdots(1+q^n)^2} = \sum_{n=0}^{\infty} V_3(n)q^n, \tag{6}$$

$$v_4(q) = 1 + \sum_{n=1}^{\infty} \frac{(-1)^n q^{2n^2}}{(1+q)(1+q^3)\cdots(1+q^{4n-3})(1+q^{4n-1})} = \sum_{n=0}^{\infty} V_4(n)q^n. \tag{7}$$

Furthermore there is the great oscillation in size of $V_i(n)$ and each has a lengthy sign change pattern that alters fairly infrequently. Thus it is reasonable to suggest that appropriate analogs of Conjectures 4–6 hold for $V_i(n)$ with $i = 2, 3, 4$.

GENERAL PROBLEMS. The point of all this is to show that there are numerous unsolved problems related to fairly simple q-series. They naturally suggest more general questions.

Problem 1. Are there any non-trivial examples of q-series which, while not equal to theta series or false theta series, nonetheless have coefficients bounded in absolute value?

Problem 2. Are there reasonable criteria to distinguish those q-series in which the coefficients in absolute value tend to infinity?

Since I can prove none of Conjectures 1–6, I have little to contribute to the solution of these more general questions. However progress on Conjectures 1–6 might allow advances on these more general questions.

It is a very easy matter to compute the power series coefficients for functions such as those in equations (1)–(7). I will be happy to supply upon request a simple BASIC program implementable on most small home computers.

This work was partially supported by National Science Foundation Grant DMS-8503324.

References

1. G. E. Andrews, On the theorems of Watson and Dragonette for Ramanujan's mock theta functions, Amer. J. Math., 88 (1966) 454–490.

2. G. E. Andrews, Enumerative proofs of certain q-identities, Glasgow Math. J., 8 (1967) 33–40.

3. G. E. Andrews, The Theory of Partitions, Encyclopedia of Mathematics and Its Applications, Vol. 2, G.-C. Rota ed., Addison-Wesley, 1976.

4. G. E. Andrews, Ramanujan's "Lost" notebook IV. Stacks and alternating parity in partitions, Adv. in Math., 53 (1984) 55–74.

5. G. E. Andrews, Ramanujan's "Lost" notebook V. Euler's partition identity, Adv. in Math. (to appear).

6. L. A. Dragonette, Some asymptotic formulae for the mock theta series of Ramanujan, Trans. Amer. Math. Soc., 72 (1952) 474–500.

7. P. Erdős, On an elementary proof of some asymptotic formulas in the theory of partitions, Ann. of Math., 43 (1942) 437–450.

8. G. Meinardus, Über Partitionen mit Differenzenbedingungen, Math. Z., 59 (1954) 388–398.

9. L. J. Rogers, On two theorems of combinatory analysis and some allied identities, Proc. London Math. Soc. (2), 16 (1916) 315–336.

10. G. N. Watson, The final problem: an account of the mock theta functions, J. London Math. Soc., 11 (1936) 55–80.

NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.

For instructions about submitting Notes for publication in this department see the inside front cover.

AN ELEMENTARY APPROACH TO THE JORDAN FORM OF A MATRIX

H. VÄLIAHO

Department of Mathematics, University of Helsinki, Hallituskatu 15, SF-00100 Helsinki 10, Finland

1. Introduction. There are numerous proofs of the existence of the Jordan normal form of a square complex matrix. In addition to the classical algebraic and geometric approaches (see, e.g., Gantmacher [4]), several elementary proofs have been presented, for example Noble [5], Filippov [1] (see also [7]), Galperin and Waksman [3], and Fletcher and Sorensen [2]. Filippov as well as Galperin and Waksman establish the existence of the Jordan form by induction, applying the theory of linear mappings. In an inductive step, Filippov reduces by one the order of every Jordan block associated with an eigenvalue of the matrix, whereas Galperin and Waksman determine these blocks entirely. Thus the inductive step of Galperin and Waksman is, in general, equivalent to several consecutive steps of Filippov, associated with the same eigenvalue. Noble's proof, like Fletcher's and Sorensen's, are given in terms of matrices. With the aid of the Schur

unitary triangularization [6], they reduce the determination of the Jordan form of a square complex matrix to the determination of the Jordan forms of nilpotent matrices. The main result of Fletcher and Sorensen is refining a strict upper triangular matrix to Jordan form by means of a sequence of similarity transformations. The pivot of Noble's proof is the construction of the Jordan strings of a general nilpotent matrix. The proofs of the main theorems of Noble and of Fletcher and Sorensen are constructive.

We shall propose an elementary inductive proof of the existence of the Jordan form. The basic idea of the induction is the same as in Galperin and Waksman [3]. The construction of Jordan strings within the inductive step bears some resemblance to Noble's [5] approach. It is, however, carried out by a more straightforward method. Contrary to [2] and [5], no preliminary unitary triangularization is needed, although a simplified version of the proof, using this triangularization, is outlined in a remark. The only nonconstructive step in the proof is the determination of the eigenvalues of a matrix.

Of the earlier elementary approaches to the Jordan form mentioned above, only Galperin and Waksman [3] includes a proof of the uniqueness of the Jordan form; Noble [5] establishes this result for the part of nilpotent matrices. We give the proof of Galperin and Waksman in a more elementary and shorter form.

If $A \in \mathbb{C}^{m \times n}$ (A is a complex $m \times n$ matrix), let $r(A)$, $\mathscr{R}(A)$ and $\mathscr{N}(A)$ stand for the rank, the column space, and the null space of A, respectively. The spectrum of a square matrix A is denoted by $\sigma(A)$, and the similarity of two square matrices A and B is indicated by $A \approx B$. The identity matrix will be denoted by I and the dimension of a vector space S by dim S. The matrix

$$J_k(\sigma) = \begin{bmatrix} \sigma & 1 & & & \\ & \cdot & \cdot & & \\ & & \cdot & \cdot & \\ & & & \cdot & 1 \\ & & & & \sigma \end{bmatrix}_{k \times k}$$

with $J_1(\sigma) = [\sigma]$ is called a *Jordan block*. A *Jordan matrix* J is a block diagonal matrix $J = J_1 \oplus \cdots \oplus J_s$, where the diagonal blocks J_i are Jordan blocks.

2. Results. We shall give an elementary proof of the following theorem:

THEOREM (Jordan form). *For any $A \in \mathbb{C}^{n \times n}$ there exists a nonsingular matrix $P \in \mathbb{C}^{n \times n}$ such that*

$$(1) \qquad P^{-1}AP = J = J_1 \oplus \cdots \oplus J_s,$$

where $J_i = J_{n_i}(\lambda_i)$, $\lambda_i \in \sigma(A)$, $i = 1, \cdots, s$. The matrix J, termed the Jordan normal form of A, is unique up to the order of the Jordan blocks.

REMARK 1. Denote the columns of P in (1) by $x_{i1}, \ldots, x_{in_i}$, $i = 1, \ldots, s$, successively. Then (1) is equivalent to

$$(2) \qquad Ax_{i1} = \lambda_i x_{i1}; \quad Ax_{ij} = \lambda_i x_{ij} + x_{i,j-1}, \quad j = 2, \ldots, n_i,$$

for $i = 1, \ldots, s$, where the vectors x_{ij} form a basis of $\mathbb{C}^n$. Any such sequence $x_{i1}, \ldots, x_{in_i}$ is called a *Jordan string*.

REMARK 2. Denote $B_i = A - \lambda_i I$; it follows from (2) that

$$x_{ij} = B_i^{n_i - j} x_{in_i}, \quad j = 1, \ldots, n_i.$$

Because, in addition, $B_i x_{i1} = 0$, we have

$$x_{i1} \in \mathscr{N}(B_i) \cap \mathscr{R}\left(B_i^{n_i - 1}\right) \quad \text{and} \quad x_{in_i} \in \mathscr{N}\left(B_i^{n_i}\right).$$

Proof of the Theorem. (*Uniqueness*; cf. [3].) Let $\lambda \in \sigma(A)$ be of algebraic multiplicity ν and let

there exist, associated with λ, q_i Jordan blocks of order i, $i = 1, \dots, n$ (set $q_i = 0$ for $i > n$). Note that for $i = 0, 1, 2, \dots$, the rank of $J_k^i(\sigma)$ equals k or $\max\{0, k - i\}$ according to whether $\sigma \neq 0$ or $\sigma = 0$. Now

$$B := A - \lambda I = P(J - \lambda I)P^{-1},$$

implying

$$r(B^i) = \sum_{j=i+1}^{n+2} (j - i)q_j + (n - \nu), \qquad i = 0, 1, \dots, n + 1.$$

From these equations we obtain uniquely

$$q_i = r(B^{i-1}) - 2r(B^i) + r(B^{i+1}), \qquad i = 1, \dots, n. \tag{3}$$

(*Existence.*) We proceed by induction on the order of A. In the inductive step, we first construct a set of vectors forming the Jordan strings associated with an eigenvalue of A and then extend this set to a basis of $\mathbb{C}^n$.

The theorem is obvious for $n = 1$. We assume it to hold for matrices of order $< n$. To establish the existence of the Jordan form of a matrix $A \in \mathbb{C}^{n \times n}$, we choose any eigenvalue λ of A, define $B = A - \lambda I$, let p be the smallest positive integer for which $r(B^{p+1}) = r(B^p) =: q_0$ and, motivated by Remark 2, define the subspaces

$$S_i = \mathcal{N}(B) \cap \mathcal{R}(B^{i-1}) = \{ y = B^{i-1}t \mid t \in \mathcal{N}(B^i) \}, \qquad i = 1, \dots, p. \tag{4}$$

Then S_i and $\mathcal{N}(B^{i-1})$, respectively, are the range and the kernel of the linear mapping

$$G_i \colon \mathcal{N}(B^i) \to \mathbb{C}^n, \quad t \mapsto y = B^{i-1}t,$$

whence

$$\dim S_i = \dim \mathcal{N}(B^i) - \dim \mathcal{N}(B^{i-1}) = r(B^{i-1}) - r(B^i).$$

We deduce that

$$\{0\} \neq S_p \subset S_{p-1} \subset \cdots \subset S_1 = \mathcal{N}(B).$$

The leading vectors y_{ij} of the Jordan strings of B, associated with the zero eigenvalue, are determined as follows: starting from a basis $y_{p1}, \dots, y_{pq_p}$ of S_p, we extend the basis of S_{i+1} sequentially for $i = p - 1, p - 2, \dots, 1$, to that of S_i by means of vectors $y_{i1}, \dots, y_{iq_i}$ (if $S_i = S_{i+1}$, this sequence is empty). Then, by (4), there exist vectors $t_{ij} \in \mathcal{N}(B^i)$ such that

$$y_{ij} = B^{i-1}t_{ij}, \quad j = 1, \dots, q_i, \quad i = 1, \dots, p.$$

We shall show that the vectors

$$B^{i-1}t_{ij}, B^{i-2}t_{ij}, \dots, Bt_{ij}, t_{ij}, \quad j = 1, \dots, q_i, \quad i = 1, \dots, p \tag{5}$$

together with any basis

$$Z = [z_1, \dots, z_{q_0}] \tag{6}$$

of $\mathcal{R}(B^p)$ form a basis of $\mathbb{C}^n$. First, the number of these vectors is

$$q_0 + \sum_{i=1}^{p} iq_i = q_0 + \sum_{i=1}^{p} \sum_{j=1}^{i} q_i = q_0 + \sum_{j=1}^{p} \sum_{i=j}^{p} q_i = q_0 + \sum_{j=1}^{p} \dim S_j = n.$$

To establish the linear independence of the vectors included in (5–6), let

$$f := Z\beta + \sum_{i=1}^{p} \sum_{j=1}^{q_i} \sum_{k=1}^{i} \alpha_{ijk} B^{i-k} t_{ij} = 0, \tag{7}$$

where β is a q_0-vector and the α_{ijk} are scalars, implying $B^p f = B^p(Z\beta) = 0$. The linear mapping

$$G\colon \mathscr{R}(B^p) \to \mathscr{R}(B^{p+1}) = \mathscr{R}(B^p), \quad t \mapsto Bt,$$

is surjective. To see this, note that any $y \in \mathscr{R}(B^{p+1})$ can be expressed in the form $y = B^{p+1}t = B(B^p t)$ with $t \in \mathbb{C}^n$, whence $y = Bu$ with $u = B^p t \in \mathscr{R}(B^p)$. Because G is surjective, it is an isomorphism, and so is G^p. Thus $B^p(Z\beta) = 0 \Rightarrow Z\beta = 0 \Rightarrow \beta = 0$. The equation (7) reduces to

$$f_1 := \sum_{k=1}^{p} \sum_{i=k}^{p} \sum_{j=1}^{q_i} \alpha_{ijk} B^{i-k} t_{ij} = 0. \tag{8}$$

To show that the α_{ijk} in (8) are equal to zero, we apply induction on (decreasing) k. From $B^{p-1}f_1 = 0$ we deduce easily that $\alpha_{pjp} = 0$, $j = 1,\ldots,q_p$. Then, assuming $\alpha_{ijk} = 0$ for $j = 1,\ldots,q_i$, $i = k,\ldots,p$, $k = h+1,\ldots,p$, we obtain

$$B^{h-1}f_1 = \sum_{i=h}^{p} \sum_{j=1}^{q_i} \alpha_{ijh} B^{i-1} t_{ij} = \sum_{i=h}^{p} \sum_{j=1}^{q_i} \alpha_{ijh} y_{ij} = 0,$$

implying $\alpha_{ijh} = 0$ for $j = 1,\ldots,q_i$, $i = h,\ldots,p$. This completes the proof of the linear independence of the vectors included in (5–6).

Because G is an isomorphism, we have $BZ = ZC$, where C (of order q_0) is nonsingular. If X is the matrix consisting of the vectors (5) in the indicated order, then

$$B[Z, X] = [Z, X](C \oplus K_1),$$

where K_1 is a Jordan matrix with zero diagonal elements, the blocks of which correspond to the strings (5). So $B \approx C \oplus K_1$. According to the induction hypothesis, C has a Jordan form, say K_2. But then

$$B \approx K_2 \oplus K_1 =: K,$$

and $A = B + \lambda I$ has the Jordan form $J := K + \lambda I$.

REMARK 3. In the proof of the existence of the Jordan form, the Jordan strings associated with a given eigenvalue of A are constructed in an inductive step. Thus the proof yields a means of determining the Jordan form and the Jordan strings of a matrix, provided that its eigenvalues are given (if only the Jordan form is needed, equation (3) can be applied).

REMARK 4. It is not difficult to reduce the determination of the Jordan form of a square complex matrix to the determination of the Jordan forms of nilpotent matrices (see, e.g., [5]). It should be noted that the proof of the Theorem is especially simple in the case of a nilpotent matrix $A \neq 0$: $\lambda = 0$, $B = A$, p is the index of A, $q_0 = 0$, and Z, C and β are void. The proof is complete in one step; no induction is needed.

References

1. A. F. Filippov, A short proof of the theorem on the reduction of a matrix to Jordan form, Moscow Univ. Math. Bull., 26 (1971) 70–71.

2. R. Fletcher and D. C. Sorensen, An algorithmic derivation of the Jordan canonical form, this MONTHLY, 90 (1983) 12–16.

3. A. Galperin and Z. Waksman, An elementary approach to Jordan theory, this MONTHLY, 87 (1981) 728–732.

4. F. R. Gantmacher, The Theory of Matrices, Vol. 1, Chelsea, New York, 1959.

5. B. Noble, Applied Linear Algebra, Prentice-Hall, Englewood Cliffs, NJ, 1969.

6. I. Schur, Über die characteristischen Wurzeln einer linearen Substitution mit einer Anwendung auf die Theorie der Integralgleichungen, Math. Ann., 66 (1909) 488–510.

7. G. Strang, Linear Algebra and Its Applications, 2nd ed., Academic Press, New York, 1980.

His name was virtually synonymous with topology for a while. (See p. 723.)

Proof. Let P be the set of all polynomials $g(x)$ with real coefficients such that $g(0), g(c), g'(0), g'(c), \ldots, g^{(k)}(0), g^{(k)}(c), \ldots$ are all integers.

CLAIM 1. *If $g(x)$ is in P, then $\int_0^c f(x)g(x)\,dx$ is an integer.*

Proof. Successive integrations by parts give

$$\int_0^c f(x)g(x)\,dx = \left[f_1 \cdot g - f_2 \cdot g' + f_3 \cdot g'' - \cdots + (-1)^d f_{d+1} \cdot g^{(d)}\right]_0^c,$$

where d is the degree of $g(x)$. This proves the claim.

We will also need the following easy fact.

(1) If $g(x)$ and $h(x)$ are in P, then so is $g(x)h(x)$.

Now assume that c is rational, and write $c = m/n$, where m, n are positive integers. Then one verifies:

(2) $m - 2nx$ is in P.

Let $g_k(x) = x^k(m - nx)^k/k!$ for $k = 0, 1, 2, \ldots$.

CLAIM 2. *$g_k(x)$ is in P for all k.*

Proof. Induction on k: $g_0(x) = 1$ is an element of P. For $k \geqslant 1$,

$$g_k'(x) = g_{k-1}(x)(m - 2nx).$$

By induction, g_{k-1} is in P, by (2) $m - 2nx$ is in P, and thus by (1) g_k' is in P. Since also $g_k(0)$ and $g_k(c)$ are 0, we have that g_k is in P.

Observe that $g_k(x) > 0$ on $(0, c)$, a property shared by $f(x)$, so that $\int_0^c f(x)g_k(x)\,dx > 0$. By Claim 1, the integral is also an integer; therefore

(3) $\int_0^c f(x)g_k(x)\,dx \geqslant 1$ for all k.

Let M be the maximum for $x(m - nx)$ on $[0, c]$, and L that for $f(x)$, then

$$\int_0^c f(x)g_k(x)\,dx \leqslant \int_0^c L \cdot \frac{M^k}{k!}\,dx = c \cdot L \cdot \frac{M^k}{k!}.$$

But $\lim_{k\to\infty} M^k/k! = 0$, contradicting (3). We are forced to conclude that c is irrational.

To prove the statement (a) mentioned at the beginning, observe that if $\cos(r)$ and $\sin(r)$ are rational, so are $\cos(|r|)$ and $\sin(|r|)$, and thus we can find a positive integer n such that $n \cdot \sin(|r|)$ and $n \cdot \cos(|r|)$ are integers. Apply the theorem, with $c = |r|$ and $f(x) = n \cdot \sin(x)$, to conclude that $|r|$ is irrational, hence that r is irrational.

To prove (b), observe that $r > 1$ without loss of generality, so that $\ln(r) > 0$. Write $r = m/n$ for some positive integers m, n, and apply the theorem with $c = \ln(r)$ and $f(x) = n \cdot e^x$.

Reference

1. I. Niven, A simple proof of the irrationality of π, Bull. Amer. Math. Soc., 53 (1947) 509.

ANSWER TO PHOTO ON PAGE 715

Kazimir Kuratowski.

AN APPLICATION OF FINITE DIFFERENCE ALGEBRA

BEDA CHAN
Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 1A1

In this note, finite difference algebra is used to prove equalities motivated by a counting process (1) and by order statistics (2). In standard texts, the special cases (3) and (4) are often stated and are proved by integration by parts or by taking the derivative with respect to c. See, for example, Feller [**2**, p. 173, Problems 46, 45]. See Chao [**1**] for an alternative proof of (3) and (4).

Let $\{N(t) | t \geq 0\}$ be a nonstationary Poisson process with instantaneous rate $\lambda(t)$. Then for $0 \leq c$,

$$\int_0^c \Pr\{N(t) = r-1\} \lambda(t)\, dt = \sum_{k=r}^{\infty} \Pr\{N(c) = k\} \tag{1}$$

says that the probability of the rth count occurring in $[0, c]$ equals the probability of having had r or more counts at c.

Let n independent observations be made on a random variable, with distribution $F(x)$ and density $f(x)$, $a \leq x \leq b$. Then for $a \leq c \leq b$,

$$\int_a^c n\binom{n-1}{r-1} F(x)^{r-1} f(x)(1-F(x))^{n-r}\, dx = \sum_{k=r}^{n} \binom{n}{k} F(c)^k (1-F(c))^{n-k} \tag{2}$$

says that the probability that the rth smallest observation be no larger than c equals the probability that r or more observations be no larger than c.

In the special case of a Poisson process, (1) becomes

$$\int_0^c \frac{\lambda^r}{\Gamma(r)} t^{r-1} e^{-\lambda t}\, dt = \sum_{k=r}^{\infty} e^{-\lambda c}(\lambda c)^k / k!; \tag{3}$$

in the special case of order statistics for uniform distribution on $[0,1]$, (2) becomes

$$\int_0^c \frac{\Gamma(n+1)}{\Gamma(r)\Gamma(n-r+1)} x^{r-1}(1-x)^{n-r}\, dx = \sum_{k=r}^{n} \binom{n}{k} c^k (1-c)^{n-k}. \tag{4}$$

The key fact from finite difference algebra needed here is that if $\psi(k) = \nabla\phi(k) = \phi(k) - \phi(k-1)$, then

$$\sum_m^n \psi(k) = [\phi(k)]_{m-1}^n = \phi(n) - \phi(m-1). \tag{5}$$

For the proof of (1), observe that

$$\begin{aligned} \frac{d}{dt} \Pr\{N(t) = k\} &= (\Pr\{N(t) = k-1\} - \Pr\{N(t) = k\})\lambda(t) \\ &= -\nabla \Pr\{N(t) = k\} \lambda(t). \end{aligned}$$

Thus,

$$\Pr\{N(c) = k\} = -\nabla \int_0^c \Pr\{N(t) = k\} \lambda(t)\, dt,$$

and by (5),

$$\begin{aligned} \sum_{k=r}^{\infty} \Pr\{N(c) = k\} &= -\left[\int_0^c \Pr\{N(t) = k\}\lambda(t)\, dt\right]_{r-1}^{\infty} \\ &= \int_0^c \Pr\{N(t) = r-1\}\lambda(t)\, dt. \end{aligned}$$

For the proof of (2), observe that

$$\begin{aligned}
&\frac{d}{dx}\binom{n}{k}F(x)^k(1-F(x))^{n-k}\\
&\quad=\left(\binom{n-1}{k-1}F(x)^{k-1}(1-F(x))^{n-k}-\binom{n-1}{k}F(x)^k(1-F(x))^{n-1-k}\right)nf(x)\\
&\quad=-\nabla\binom{n-1}{k}F(x)^k(1-F(x))^{n-1-k}nf(x).
\end{aligned}$$

Thus

$$\binom{n}{k}F(c)^k(1-F(c))^{n-k}=-\nabla\int_a^c\binom{n-1}{k}F(x)^k(1-F(x))^{n-1-k}nf(x)\,dx,$$

and by (5),

$$\begin{aligned}
\sum_{k=r}^{n}\binom{n}{k}F(c)^k(1-F(c))^{n-k}&=-\left[\int_a^c\binom{n-1}{k}F(x)^k(1-F(x))^{n-1-k}nf(x)\,dx\right]_{r-1}^{n}\\
&=\int_a^c\binom{n-1}{r-1}F(x)^{r-1}(1-F(x))^{n-r}nf(x)\,dx.
\end{aligned}$$

References

1. A. Chao, Another approach to incomplete integrals, The American Statistician, 36 (1982) 48.

2. W. Feller, An Introduction to Probability Theory and Its Applications, Vol 1, 3rd ed., Wiley, New York, 1968.

FINITE LINEAR GROUPS, THE COMMODORE 64, EULER AND SYLVESTER

DAVID BENSON, WALTER FEIT, AND ROGER HOWE

Department of Mathematics, Yale University, New Haven, CT 06520

For some examples in group theory, one of us was interested in knowing the number of conjugacy classes in $GL_n(2)$, where $GL_n(q)$ denotes the general linear group in dimension n over the finite field with q elements. A generating function was given in [3] for the number $p_n(q)$ of conjugacy classes in $GL_n(q)$:

$$1+\sum_{n=1}^{\infty}p_n(q)x^n=\prod_{i=1}^{\infty}\frac{(1-x^i)}{(1-qx^i)}. \tag{1}$$

Since he had available a Commodore 64 minicomputer, the aforementioned one of us let it compute not simply $p_n(2)$, but the full polynomials $p_n(q)$, for $n<40$. Here is a sample of the results.

$$\begin{aligned}
p_1(q)&=q-1,\\
p_2(q)&=q^2-1,\\
p_4(q)&=q^4-q,\\
p_8(q)&=q^8-q^3-q^2+q,\\
p_{16}(q)&=q^{16}-q^7-q^6-q^5+2q^3+q^2-q,\\
p_{32}(q)&=q^{32}-q^{15}-q^{14}-q^{13}-q^{12}-q^{11}-q^{10}+q^9\\
&\quad+2q^8+4q^7+3q^6-4q^4-3q^3+q^2+q.
\end{aligned} \tag{2}$$

Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T: Textbook	P: Professional Reading	1-4: Semesters
C: Computer Software	L: Undergraduate Library	** : Special Emphasis
S: Supplementary Reading	13: Grade Level	?? : Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the Monthly.

General, S*, P*, L*. The Second Creation: Makers of the Revolution in Twentieth-Century Physics. Robert P. Crease, Charles C. Mann. Macmillan, 1986, xi + 480 pp, $25. [ISBN: 0-02-521440-3] An absorbing account of the motivation, methods, personalities, and compelling intellectual curiosity that is driving twentieth-century physicists towards the elusive unified theory of energy and matter. Extensive quotations from interviews enrich the authors' lucid exposition, weaving human motives with scientific logic in a rich tapestry of discovery and understanding. LAS

General, P. Summary Report 1984: Doctorate Recipients from United States Universities. Susan L. Coyle, Peter D. Syverson. National Academy Pr, 1986, vi + 49 pp, (P). A summary report on trends, ethnic diversity, baccalaureate sources, citizenships and fields of the 31,253 1984 U.S. doctorate recipients. Features a study of the "productivity" of undergraduate institutions by comparing 1984 doctorates with 1974 baccalaureate degrees. LAS

General, L, P*.** The Influence of Computers and Informatics on Mathematics and Its Teaching. Ed: A.G. Howson, J.-P. Kahane. ICMI Study Ser: Strasbourg 1985. Cambridge U Pr, 1986, vii + 155 pp, $12.95 (P). [ISBN: 0-521-31189-6] Formal report of a symposium organized at Strasbourg in March 1985 by the International Commission on Mathematical Instruction (ICMI). First in a series of studies by ICMI, this volume is designed to stimulate further world-wide discussion of the expanding influence of computers on mathematical research and teaching, especially at the university level. Contains a 40-page paper summarizing important themes from the Strasbourg symposium, supported by 11 papers selected and revised from the 49 contained in the Supporting Papers previously published by IREM, Université Louis Pasteur. Authors include Atiyah, de Bruijn, Davenport, Stoutemeyer, and Burkhardt; topics include symbolic algebra, discrete mathematics, calculus and the computer, and computer graphics. LAS

Elementary, T(13: 1). Electronics Math, Second Edition. Bill Deem. Prentice-Hall, 1986, xiv + 607 pp, $32.95. [ISBN: 0-13-252321-3-01] Basic arithmetic, algebra, and trigonometry to solve electronics applications. Problems revolve around AC and DC circuit analysis. One-color printing provides little variation in format or presentation. RD

Elementary, T(13-16: 1). Mathematics: Its Power and Utility, Second Edition. Karl J. Smith. Brooks/Cole, 1986, xiv + 495 pp, $31.50. [ISBN: 0-534-05268-1] Gives student "who has previously not been successful with mathematics a fresh amd innovative approach to arithmetic, beginning algebra, and geometry." Preface gives no indication of changes made in this edition (First Edition, TR, June-July 1983). JNC

Precalculus, T(13: 1). Precalculus Mathematics: A Functional Approach, Third Edition. Karl J. Smith. Brooks/Cole, 1986, xvi + 522 pp, $32.50. [ISBN: 0-534-05232-0] Unifying concepts: functions and graphs. Polynomial and rational functions, exponential and logarithmic functions, trigonometric functions. Also, chapters on analytic geometry, matrices (including a new section on linear programming), induction, series and sequences. Large problem sets, graded for difficulty. New in this edition: second chapter on analytic geometry, vectors (two and three dimensions), polar coordinates, parametric equations, some reorganization of topics (e.g., complex numbers are integrated into the text), reworked problem sets. LCL

Precalculus, T(13: 1). College Algebra with Applications, Second Edition. M.A. Munem, D.J. Foulis. Worth, 1986, xiii + 460 pp, $27.95. [ISBN: 0-87901-285-4] Only relatively minor changes and additions to the First Edition (TR, November 1982) including more on complex numbers, zeros of polynomials, and analytic geometry. JS

Precalculus, T(13: 1). Algebra and Trigonometry with Applications, Second Edition. M.A. Munem, D.J. Foulis. Worth, 1986, xiv + 662 pp, $29.95. [ISBN: 0-87901-281-1] Only relatively minor changes and additions to the First Edition (TR, November 1982); most notable is expanded analytic geometry. JS

Precalculus, T(13: 1). College Algebra and Trigonometry with Applications. Thomas Koshy. McGraw-Hill, 1986, xiv + 666 pp, $31.95. [ISBN: 0-07-035471-5] Standard precalculus material. Extensive exercise sets--noncalculator and calculator varieties. Good supply of word problems and applications. Warning symbols to alert students to frequently made errors and to notions of importance in calculus. Helpful, if heeded, down-to-earth study suggestions and tips on use of hand-held calculator. JK

Precalculus, T(13: 1, 2). Precalculus: Functions and Graphs, Fourth Edition. M.A. Munem, J.P. Yizze. Worth, 1985, xi + 557 pp, $27.95. [ISBN: 0-8791-258-7] Covers topics in algebra and trigonometry needed for calculus and linear algebra. Assumes geometry and intermediate algebra. Emphasizes graphing techniques. Calculator exercises and suggestions on calculator use. Changes from the Third Edition include more applications, expanded coverage of graphing, polar coordinates, matrices, sequences. KS

Precalculus, T(13: 1). Technical Algebra with Applications. C.E. Goodson, S.L. Miertschin. Wiley, 1985, xiii + 572 pp, $26.95. [ISBN: 0-471-08241-4] Text for students in technical and pre-engineering programs. Covers linear, quadratic and higher degree equations, logarithmic and exponential functions, plane analytic geometry, sequences and series, probability, statistics, matrices. Optional sections on engineering and business applications. Appendices on calculator use and BASIC. KS

Education, P*, L. Interactive Computer Programs for Education: Philosophy, Techniques, and Examples. Jay Nievergelt, Andrea Ventura, Hans Hinterberger. Addison-Wesley, 1986, viii + 190 pp, (P). [ISBN: 0-201-11129-2] Presents guidelines and techniques for developing interactive computer-based demonstrations. Also includes some examples. Of particular interest to secondary school educators. AO

History, S, P, L.** The Rise of Statistical Thinking 1820-1900. Theodore M. Porter. Princeton U Pr, 1986, xii + 333 pp, $35. [ISBN: 0-691-08416-5] A detailed account of the roots of modern statistics in nineteenth century science--of "the doctrine that order is to be found in large numbers." Contributions of Quetelet, Galton, Boltzmann, Maxwell, Edgeworth and Pearson play a major role as do the interdependent contributions of the social ("statist") sciences and the natural (biological and physical) sciences. LAS

Logic, T(18: 1), P. Self-Reference and Modal Logic. C. Smoryński. Universitext. Springer-Verlag, 1985, xii + 333 pp, $29.80 (P). [ISBN: 0-387-96209-3] Gödel's Incompleteness Theorem involves logical statements which refer to their own provability. This text for graduate seminar or self-study explores self-referential statements using modal techniques in order to understand and apply self-reference. KS

Logic, P. Lecture Notes in Mathematics-1103: Models and Sets. Ed: G.H. Müller, M.M. Richter. Springer-Verlag, 1984, viii + 484 pp, $22.50 (P) [ISBN: 0-387-13900-1]; Lecture Notes in Mathematics-1104: Computation and Proof Theory. Ed: M.M. Richter, et al. 1984, viii + 475 pp, $22.50 (P). [ISBN: 0-387-13901-X] Proceedings of the Logic Colloquium held July 18-23, 1983 in Aachen, Federal Republic of Germany. LNM-1103 contains twenty-one research papers on model theory of algebraic structures, nonstandard analysis, admissible ordinals, Boolean algebra; LNM-1104 contains an additional twenty-one research papers on recursion theory, applications of logic to computer science, and proof theory. KS

Logic, P. Lecture Notes in Mathematics-1130: Methods in Mathematical Logic. Ed: C.A. Di Prisco. Springer-Verlag, 1985, vii + 407 pp, $25.80 (P). [ISBN: 0-387-15236-9] Proceedings of sixth Latin American Symposium on Mathematical Logic held in Caracas, Venezuela, August 1-6, 1983. Diverse collection of eighteen papers including survey of applications of model theory to real algebraic geometry by Dickmann and introduction to categorical logic by Makkai. KS

Logic, T(18: 1), P. Recursive Aspects of Descriptive Set Theory. Richard Mansfield, Galen Weitkamp. Logic Guides, V. 11. Oxford U Pr, 1985, vii + 144 pp, $19.95. [ISBN: 0-19-503602-6] Descriptive set theory is the study of mathematical properties of subsets of the real numbers definable by some types of logical formulas. This book provides an introduction to the subject designed to prepare the reader to understand Friedman's work on assumptions required to prove mathematical statements. Prerequisites are elementary recursion theory and set theory. KS

Logic, T(13: 1), L. The Logic Basis for Computer Programming, Volume I: Deductive Reasoning. Zohar Manna, Richard Waldinger. Addison-Wesley, 1985, xii + 618 pp. [ISBN: 0-201-18260-2] Text for introductory logic course for computer science students. First part covers elementary propositional and predicate logic; emphasizes methods for proving validity. Second part presents natural numbers, strings, trees, lists, sets, tuples as logical theories. No college-level mathematics or programming experience assumed, but some mathematical maturity needed. Authors believe this material should replace calculus as requirement for computer science majors. KS

Logic, T(13-14: 1), S, L. Logic: A Computer Approach. Morton L. Schagrin, William J. Rapaport, Randall R. Dipert. McGraw-Hill, 1985, xviii + 347 pp, $19.95 (P). [ISBN: 0-07-055131-6] Clearly written, leisurely-paced text for introductory logic course, particularly for computer science students. Covers basics of propositional and predicate calculus. Appendices on circuit design and Turing machines. Uses pseudolanguage to present algorithms for tasks including truth table generation and proof construction. Routine exercises. KS

Logic, P. Recursion Theory. Ed: Anil Nerode, Richard A. Shore. Proc. of Symp. in Pure Math., V. 42. AMS, 1985, vii + 528 pp, $60. [ISBN: 0-8218-1447-8] Twenty-seven papers based on short courses and invited lectures presented at AMS-ASL Summer Research Institute held at Cornell University, June 28-July 16, 1982. Divided into five sections: classical recursion theory, generalized recursion theory, fine structure and descriptive set theory, effective mathematics, foundations and complexity theory. KS

Graph Theory, P. Graphs and Applications: Proceedings of the First Colorado Symposium on Graph Theory. Ed: Frank Harary, John S. Maybee. Wiley, 1985, xv + 347 pp, $44.95. [ISBN: 0-471-88772-2] Twenty-three papers from a symposium held at the University of Colorado, Boulder, October 15-16, 1986. Includes expository article on Petersen graph. Some papers mention open problems. KS

Combinatorics, S(14-17), L*. Problems in Combinatorics and Graph Theory. Ioan Tomescu. Transl: Robert A. Melter. Ser. in Disc. Math. Wiley, 1985, xvii + 335 pp, $31.95. [ISBN: 0-471-80155-0] Revision and translation of 1981 Romanian edition. Over three hundred problems divided into fourteen topics. Complete solutions. Great range of difficulty; some suitable for introductory courses, others from research journals. Author uses book to prepare high school students for International Mathematical Olympiad. KS

Discrete Mathematics, T(13-15: 1). Discrete Mathematics for Computer Scientists and Mathematicians, Second Edition. Joe L. Mott, Abraham Kandel, Theodore P. Baker. Reston, 1986, xiv + 751 pp, $31.95. [ISBN: 0-8359-1391-0] There are two major changes from the First Edition (TR, May 1984): a new chapter on network flows and matching, and a separate chapter on fuzzy sets and imprecision. JS

Number Theory, T(16-17: 1), S, P, L. Introduction to Arithmetical Functions. Paul J. McCarthy. Universitext. Springer-Verlag, 1986, vii + 365 pp, $35.50 (P). [ISBN: 0-387-96262-X] An introduction to the theory of arithmetical functions which assumes some mathematical sophistication and an introduction to elementary number theory. A solid text with a contemporary flavor. An outstanding collection of over 400 exercises is included. CEC

Topological Groups, P. Lie Algebras and Related Topics. Ed: D.J. Britten, F.W. Lemire, R.V. Moody. Conf. Proc., V. 5. AMS, 1984, vii + 382 pp, $44 (P). [ISBN: 0-8218-6009-7] Lecture courses and research articles and announcements from a summer seminar at the University of Windsor, Canada, in 1984. BC

Topological Groups, P. Representations of Lie Groups and Lie Algebras. Ed: A.A. Kirillov. Akademiai Kiado, 1985, 225 pp, $24. [ISBN: 963-05-3542-4] Seven introductory lectures on representation theory, from a Summer School in Budapest. Accessible to non-specialists. BC

Algebra, T*(16-17: 1, 2), S, L. Abstract Algebra. Gary D. Crown, Maureen H. Fenrick, Robert J. Valenza. Pure & Appl. Math., V. 99. Dekker, 1986, vi + 403 pp, $32.50. [ISBN: 0-8247-7456-6] This text covers the fundamental concepts of abstract algebra at a level appropriate for first-year graduate students. A clearly written, rigorous exposition which is supplemented with a wealth of examples and lots of exercises at different levels of difficulty. CEC

Algebra, T.** Abstract Algebra. I.N. Herstein. Macmillan, 1986, xiii + 289 pp. [ISBN: 0-02-353820-1] This short and sweet book is related to the author's famous Algebra only by its common topic. It provides a treatment of groups, rings, and fields which is very insightful yet cognizant of today's less well (and differently) prepared students. It remains a treatment of core mathematical ideas, treated for their beauty and power without apology (except in Hardy's sense) for the fact that many students find these ideas quite abstract. Three levels of problems will help an instructor meet his students needs. JAS

Algebra, S(17-18), P. Bilinear Forms and Orderings on Commutative Rings. Murray A. Marshall. Papers in Pure & Appl. Math., No. 71. Queen's U, 1985, 175 pp, (P). Lecture notes on the interplay between the Witt ring of bilinear forms over a commutative ring A and the real spectrum of A. LAS

Calculus, T(13: 1). The Power of Calculus, Fourth Edition. Kenneth L. and Mary Neil Whipkey. Wiley, 1986, xiii + 474 pp, $30.95. [ISBN: 0-471-06382-7] This edition (First Edition, TR, August-September 1972; Second Edition, TR, December 1975; Third Edition, TR, June-July 1979) adds new "real life" problems, graphs, and illustrations to a rewriting of the text which reduces the mathematical level and notation. JNC

Calculus, T(13), S. BASIC Computing for Calculus. Ronald I. Rothenberg. McGraw-Hill, 1985, xv + 309 pp, $37.95 (P). [ISBN: 0-07-054011-X] A collection of BASIC microcomputer programs and exercises, together with expository material reviewing rudiments of elementary calculus. Most programs illustrate calculus ideas numerically or compute numerical approximations (e.g., trapezoid rule). First two chapters introduce BASIC language. Only last chapter, which emphasizes Macintosh graphics, is hardware-specific. PZ

Calculus, T*(13-14: 1-3). Calculus, One and Several Variables with Analytic Geometry, Fifth Edition. S.L. Salas, Einar Hille, John T. Anderson. Wiley, 1986, xv + 1129 pp, $39.95. [ISBN: 0-471-87549-X] This edition features numerous changes (First Edition, TR, June-July 1971; Second Edition, TR, October 1974; Third Edition, TR, June-July 1978 and December 1978; Fourth Edition, TR, June-July 1982). Noteworthy are an early introduction of trigonometry which allows early coverage but also permits delay without disruption, better balanced exercise sets, more motivational pieces, fuller treatment of numerical techniques, a middle-of-the-book chapter on modelling and differential equa-

tions, revision of some key sections and shuffling of chapters. Still a solid accessible-to-students treatment. Available in a single complete volume or in two parts. JK

Real Analysis, T(16-17: 1), P. Lectures on Constructive Mathematical Analysis. B.A. Kushner. Transl. of Math. Mono., V. 60. AMS, 1984, v + 346 pp, $95. [ISBN: 0-8218-4513-6] Translation of 1973 Russian monograph provides an introduction to algorithmic aspects of analysis with emphasis on undecidability results. Assumes familiarity with logical notation and elementary real analysis. KS

Real Analysis, P. Lecture Notes in Mathematics-1170: Real Functions. Brian S. Thomson. Springer-Verlag, 1985, vii + 229 pp, $14.40 (P). [ISBN: 0-387-16058-2] Using "local systems" (a structure similar to that of a filter in topology) as a framework, the text surveys the topics of real cluster sets and generalized notions of limit, derivative, and continuity for real functions. Background in basic real analysis (measure, density, category) required. Extensive bibliography. BH

Complex Analysis, P. Analytic Functions of One Complex Variable. Ed: Chung-chun Yang, Chi-tai Chuang. Contemp. Math., V. 48. AMS, 1985, x + 254 pp, $26 (P). [ISBN: 0-8218-5050-4] Collection of expository and survey articles in one complex variable by Chinese mathematicians. BH

Complex Analysis, T(18), S, P. Applied and Computational Complex Analysis, Volume 3: Discrete Fourier Analysis--Cauchy Integrals--Construction of Conformal Maps--Univalent Functions. Peter Henrici. Wiley, 1986, xiii + 637 pp, $59.95. [ISBN: 0-471-08703-3] Written from the viewpoint that a problem is not solved unless one is able to "provide an algorithm (implementable on computers) for constructing solutions." Problems and thorough notes at end of each section. Very extensive bibliography. BH

Complex Analysis, P. Lecture Notes in Mathematics-1171: Polynômes Orthogonaux et Applications. Ed: C. Brezinski, et al. Springer-Verlag, 1985, xxxvii + 584 pp, $42 (P). [ISBN: 0-387-16059-0] Proceedings of the Laguerre Symposium held at Bar-le-Duc, October 15-18, 1984. JAS

Differential Equations, T*(14-15: 1), L. A Second Course in Elementary Differential Equations. Paul Waltman. Academic Pr, 1986, xi + 259 pp, $27. [ISBN: 0-12-73[illegible] [illegible] for such a course, whether it follows an introduction in the calculus sequence or a regular first course. Four chapters, on linear systems (which uses the Putzer algorithm rather than the Jordan Canonical Form theorem); two-dimensional systems (includes Liapunov theory and the Poincaré-Bendixson theorem); existence theory; boundary value problems. Includes nontrivial applications. Nice mixture of computational, intuitive, and rigorous approaches. Well worth a good look. DFA

Partial Differential Equations, S(17-18), P. Free and Moving Boundary Problems. John Crank. Clarendon Pr, 1984, x + 425 pp, $64. [ISBN: 0-19-853357-8] A boundary value problem is a differential equation whose solutions must satisfy side conditions along the boundary of a domain. Moving boundary value problems are associated with time dependent problems where the position of the boundary varies with time. This book describes methods of solution for these problems and includes results from computer studies of these problems in two- and three-dimensions. AM

Partial Differential Equations, P. Boundary Integral Equation Methods in Eigenvalue Problems of Elastodynamics and Thin Plates. Michihiro Kitahara. Stud. in Appl. Mech., V. 10. Elsevier Science, 1985, viii + 281 pp, $66.75. [ISBN: 0-444-42447-4] Boundary integral equation (BIE) methods are used in the analysis of partial differential equations, particularly wave propagation problems. Discusses the application of BIE methods to the analysis of eigenvalue problems arising in elastodynamics and thin plates. Includes numerical techniques and contains bibliographic references to current literature. AM

Partial Differential Equations, P. Lecture Notes in Mathematics-1134: Weighted Energy Methods in Fluid Dynamics and Elasticity. Giovanni P. Galdi, Salvatore Rionero. Springer-Verlag, 1985, vii + 126 pp, $9.50 (P). [ISBN: 0-387-15645-3] Weighted energy methods refer to a collection of techniques using auxiliary weight functions in the estimation of solutions to partial differential equations. This text demonstrates the use of these techniques to solve several problems arising in fluid dynamics and elasticity. BH

Numerical Analysis, T(13: 1), S, L. Basic Numerical Methods: An Introduction to Numerical Mathematics on a Microcomputer. R.E. Scraton. Edward Arnold, 1984, viii + 92 pp, $13.95 (P). [ISBN: 0-7131-3521-2] Elementary numerical methods for solution of nonlinear equations, linear systems, ordinary differential equations, and numerical integration. Numerical stability and ill-conditioning are briefly mentioned. BASIC programming language assumed. SM

Numerical Analysis, P*. Solution of Partial Differential Equations on Vector and Parallel Computers. James M. Ortega, Robert G. Voigt. SIAM, 1985, iii + 96 pp, $12 (P). [ISBN: 0-89871-055-3] A survey of the state-of-the-art. Discusses both direct and iterative methods for elliptic equations as well as both explicit and implicit methods for initial-boundary value problems. AO

Numerical Analysis, T(16-17: 1, 2), S. Numerical Methods for Scientific and Engineering Computation. M.K. Jain, S.R.K. Iyengar, R.K. Jain. Halsted Pr, 1985, viii + 406 pp. [ISBN: 0-470-20143-6] Practical approach to numerical methods, emphasis on computation, comparison of implementations, error analysis, method choice. Standard topics (errors, direct and iterative techniques for root finding, linear systems, interpolation and approximation, numerical differential and integral ordinary differential equations, partial differential equations). Many nice, small examples, exercises; some

from BIT. RM

Numerical Analysis, T(15-16: 1, 2). Elementary Numerical Analysis. W. Allen Smith. Reston, 1986, ix + 582 pp, $33.95. [ISBN: 0-8359-1719-3] An introductory textbook covering the standard topics. Relatively few formal theorems, but numerous exercises. Sample computer programs are given for many of the methods discussed. AO

Numerical Analysis, P. Numerical Analysis of Parametrized Nonlinear Equations. Werner C. Rheinboldt. U. of Arkansas Lect. Notes in Math. Sci., V. 7. Wiley, 1986, xi + 299 pp, $34.95 (P). [ISBN: 0-471-88814-1] An expanded version of a series of lectures given at the University of Arkansas in the spring of 1983. Covers both the practical (numerical) aspects of the problem as well as the theoretical aspects (using differentiable manifolds). AO

Functional Analysis, T(16-17). Notions de Topologie: Introduction aux Espaces Fonctionnels. Claude Tisseron. Hermann, 1985, 319 pp, 130 F (P). [ISBN: 2-7056-6019-4] An introductory text on functional analysis. The necessary general topology is covered and illustrated, whenever possible, in function spaces. JD-B

Functional Analysis, P. Lecture Notes in Mathematics-1138: Actions of Discrete Amenable Groups on von Neumann Algebras. Adrian Ocneanu. Springer-Verlag, 1985, 115 pp, $9.80 (P). [ISBN: 0-387-15663-1]

Functional Analysis, P. A Local Spectral Theory for Closed Operators. Ivan Erdelyi, Wang Shengwang. London Math. Soc. Lect. Note Ser., V. 105. Cambridge U Pr, 1985, 178 pp, $22.95 (P). [ISBN: 0-521-31314-7] If T is a closed operator acting on a complex Banach space X, then the spectral decomposition problem is to express X as a finite direct sum of invariant subspaces X_i such that the spectra of the restrictions of T to the spaces X_i are contained in some specified closed sets. This text extends this problem to the case of unbounded closed operators and develops techniques for analyzing it. BH

Functional Analysis, S(18), P. Lecture Notes in Mathematics-1169: Approximation Theory in Tensor Product Spaces. W.A. Light, E.W. Cheney. Springer-Verlag, 1985, 157 pp, $12 (P). [ISBN: 0-387-16057-4] Studies approximation of multivariate functions by combinations of univariate functions in the setting of tensor products of Banach spaces. First chapter is an introduction to this setting. Also contains appendix on Bochner integral as well as an extensive bibliography. BH

Functional Analysis, S(18), P. Unbounded Non-Commutative Integration. J.P. Jurzak. Math. Physics Stud. D Reidel, 1985, xx + 191 pp, $39.50. [ISBN: 90-277-1815-6] A research-level monograph describing the relation between the theory of unbounded non-commutative integration and questions about regularity that arise in statistical mechanics. LAS

Analysis, T(16-17: 1), S, P, L. Approximation of Functions. G.G. Lorentz. Chelsea, 1986, ix + 188 pp, $14.95. [ISBN: 0-8284-0322-8] Slightly revised reissue of the 1966 original edition (TR, April 1967; Extended Review, June-July 1968). An elementary introduction to problems of approximating (mostly) real functions with functions of more tractable types: algebraic and trigonometric polynomials, rational functions, etc. Closeness of approximation, measured in various ways, is a recurring theme. Clear and readable; each chapter has notes and exercises. PZ

Analysis, S(18), P. Hardy Classes and Operator Theory. Marvin Rosenblum, James Rovnyak. Math. Mono. Oxford U Pr, 1985, xiv + 161 pp, $39.95. [ISBN: 0-19-503591-7] Relates theory of shift operators and Toeplitz operators to Hardy classes of vector and operator valued functions. Assumes basic knowledge of function theory and operator theory. Extensive bibliography. BH

Analysis, P. Lecture Notes in Mathematics-1152: Asymptotic Expansions for Pseudodifferential Operators on Bounded Domains. Harold Widom. Springer-Verlag, 1985, 150 pp, $12 (P). [ISBN: 0-387-15701-8] Two types of asymptotic expansions are studied: the "Szegö expansion" for the traces of fairly general functions of integral operators, and the "heat expansion" for scalar elliptic pseudodifferential operators of negative order. BH

Analysis, P. Lecture Notes in Mathematics-1163: Iteration Theory and its Functional Equations. Ed: R. Liedl, L. Reich, Gy. Targonski. Springer-Verlag, 1985, viii + 231 pp, $14.40 (P). [ISBN: 0-387-16067-1] The proceedings of the international symposium held at Schloss Hofen (Lochau), Austria, September 28 to October 1, 1984. Contains both traditional technical material as well as discussion of numerical "experimental" results and some philosophical discussion of cutting-edge issues. JAS

Algebraic Geometry, S(18), P. Lecture Notes in Mathematics-1173: Locally Semialgebraic Spaces. Hans Delfs, Manfred Knebusch. Springer-Verlag, 1985, xvi + 329 pp, $20.50 (P). [ISBN: 0-387-16060-4] These lecture notes contribute to a basic aspect of semialgebraic geometry--the topological phenomena of semialgebraic sets in V(R) for V a variety over a real closed field R. Bibliography included. CEC

Differential Geometry, P. Lecture Notes in Mathematics-1161: Harmonic Mappings and Minimal Immersions. Ed: E. Giusti. Springer-Verlag, 1985, vii + 285 pp, $17.60 (P). [ISBN: 0-387-16040-X] Lectures from the first 1984 session of the Centro Internationale Matematico Estivo (CIME) held at Montecatini, Italy, June 24-July 3, 1984. JAS

Algebraic Topology, S(18), P. Three-Dimensional Link Theory and Invariants of Plane Curve Singularities. David Eisenbud, Walter Neumann. Annals of Math. Stud., No. 110. Princeton U Pr, 1985, vii + 172 pp, $39.50; $13.95 (P). [ISBN: 0-691-08380-0; 0-691-08381-9] This monograph gives a new foundation for the theory of links in 3-space modelled on the modern development by Jaco, Shalen, Johannson, Thurston, et al, of the theory of 3-manifolds. CEC

Operations Research, S(16-17), L. Optimization Models for Planning and Allocation: Text and Cases in Mathematical Programming. Roy D. Shapiro. Wiley, 1984, xiv + 650 pp, $44. [ISBN: 0-471-09468-4] A collection of case studies developed at the Harvard Business School together with some background material on mathematical programming techniques and model formulation. AO

Optimization, P, L*. Combinatorial Optimization: Annotated Bibliographies. Ed: M. O'hEigeartaigh, J.K. Lenstra, A.H.G. Rinnooy Kan. Wiley, 1985, vii + 204 pp, $24.95 (P). [ISBN: 0-471-90490-2] A collection of annotated bibliographies covering several major areas within the field of combinatorial optimization. Each bibliography was compiled and annotated by a leading researcher active in the area. AO

Optimization, P. Lecture Notes in Economics and Mathematical Systems-259: Infinite Programming. Ed: E.J. Anderson, A.B. Philpott. Springer-Verlag, 1985, xiv + 244 pp, $20.50 (P). [ISBN: 0-387-15996-7] Infinite programming is the study of mathematical programming problems in which the number of variables and the number of constraints are possibly infinite. This volume consists of 18 research papers selected from those given at the Symmposium on Infinite Dimensional Linear Programming held at Cambridge, England, September 7-10, 1984. KS

Optimization, S(15-17). Advanced Mathematics for Economists: Static and Dynamic Optimization. Peter J. Lambert. Basil Blackwell, 1985, xiv + 231 pp, $14.95 (P); $45. [ISBN: 0-631-14139-1; 0-631-14138-3] The first half is a review of basic material from univariate calculus, matrix algebra, and multivariate calculus. The second half focuses on Lagrange multiplier techniques but also includes a brief introduction to the calculus of variations and dynamic programming. AO

Optimization, P. Finite Algorithms in Optimization and Data Analysis. M.R. Osborne. Wiley, 1985, xv + 383 pp, $51.95. [ISBN: 0-471-90539-9] A unified treatment of finite algorithms for optimization and data analysis problems based on active set strategies. The main problem classes treated are minimization of polyhedral convex functions and solution of convex robust estimation problems. AO

Dynamical Systems, S*, P*, L.** The Beauty of Fractals: Images of Complex Dynamical Systems. H.-O. Peitgen, P.H. Richter. Springer-Verlag, 1986, xii + 199 pp, $29.50. [ISBN: 0-387-15851-0] A stunning photographic and mathematical display of fractal geometry arising in dynamical systems, in Julia and Mandelbrot sets, in global views of Newton's method, in Volterra-Lotka models of predator-prey systems, in the Ising model of magnetism, and in the Yang-Lee theory of phase transitions. Based on the exhibit book prepared for "Frontiers of Chaos"--an exhibition sponsored by the Goethe Institute currently on tour in the United States--this attractive volume includes invited interpretive and historical comments on science, art and mathematics, as well as details about the mappings and parameters used to create the color plates. LAS

Dynamical Systems, S(17-18), P. Methods in the Qualitative Theory of Dynamical Systems in Astrophysics and Gas Dynamics. Oleg I. Bogoyavlensky. Transl: Dmitry Gokhman. Springer-Verlag, 1985, ix + 301 pp, $49. [ISBN: 0-387-13614-2] The maximally non-degenerate compactification of a dynamical system is obtained by completion of phase space by a boundary at infinity and is used to study the asymptotic behavior of trajectories for large energies. This method is used to analyze multidimensional dynamical systems, including applications in cosmology, theoretical astrophysics, and gas dynamics. AL

Dynamical Systems, T*(15-16: 1), S, P, L*. An Introduction to Chaotic Dynamical Systems. Robert L. Devaney. Benjamin/Cummings, 1986, xiv + 320 pp. [ISBN: 0-8053-1601-9] An example-oriented introduction to nonlinear dynamical systems via the study of iterates of maps in one and two dimensions (e.g., the quadratic map $F_k(x) = kx(1-x)$, the Hénon map $H(x,y) = (a-by-x^2,x)$, Julia sets of complex quadratic polynomials). These and many other simple examples reveal the definitions, theorems and especially the geometry of bifurcations, attractors, chaos, structural stability and symbolic dynamics. Requires only knowledge of standard sophomore mathematics; emphasizes mathematical methods from analysis and topology, with few hints about applictions or computer graphics. A superb text for a senior seminar. LAS

Probability, T(16: 1), L. Markov Chains, Theory and Applications. Dean L. Isaacson, Richard W. Madsen. Robert E Krieger, 1985, x + 256 pp, $43.50. [ISBN: 0-89874-834-8] Reprint of the original 1976 edition published in the Wiley Series in Probability and Mathematical Statistics (TR, October 1976). RSK

Probability, S(14-16). Probability Distributions. V. Rothschild, N. Logothetis. Wiley, 1986, 70 pp, $7.95 (P). [ISBN: 0-471-83814-4] Contains graphs, properties, applications, and relationships with other distributions for thirty of the most common probability distributions. RSK

Probability, T(15-16: 1). Probability Theory and Applications. Enders A. Robinson. Intern Human Res Develop Corp (137 Newbury St, Boston, MA 02116), 1985, ix + 420 pp, $39. [ISBN: 90-277-2025-8] Undergraduate text in probability with calculus prerequisite; some prior knowledge of elementary probability suggested. Standard topics well treated; good selection of solved problems and examples. Exercises follow each section; appendix has 174 more. Companion computer laboratory manual

in preparation. KK

Statistics, T(14-15: 1, 2), P. Nonparametric Methods for Quantitative Analysis, Second Edition. Jean Dickinson Gibbons. Ser. in Math. & Management Sci., V. 2. American Sciences Pr, 1985, xv + 481 pp, $34.50 (P). [ISBN: 0-935950-09-5] Well written elementary handbook of nonparametric methods accessible to beginning students in statistics. Useful as reference for researchers. Addition of chapter on binomial selection procedures (with tables) is main difference from First Edition (1976). Contains 20 tables (a few could be dropped) and good set of about 180 exercises. KK

Statistics, P. Statistical Analysis of Measurement Errors. John L. Jaech. Exxon Mono. Wiley, 1985, xxiii + 293 pp, $34.95. [ISBN: 0-471-82731-2] Develops a general methodology, based on maximum likelihood, for estimation and inference when dealing with the situation where a number of items are each measured for the same characteristic by at least two measurement methods. RSK

Statistics, T(17-18: 1), P*. Analysis of Experiments with Missing Data. Yadolah Dodge. Wiley, 1985, xvii + 499 pp, $39.95. [ISBN: 0-471-88736-6] In the Wiley Series in Probability and Mathematical Statistics. Presents recently developed theories and techniques for analyzing designed experiments with missing data. Includes FORTRAN programs with sample output. Good sets of references. No exercises. RSK

Statistics, P*. Bandit Problems: Sequential Allocation of Experiments. Donald A. Berry, Bert Fristedt. Chapman & Hall, 1985, viii + 275 pp, $25. [ISBN: 0-412-24810-7] Comprehensive treatment of the problem in statistical decision theory of making sequential selections from two or more stochastic processes in order to maximize an expected payoff. Excellent annotated bibliography. RSK

Statistics, P*. Modern Statistical Methods in Digital Simulation. Ed: Pandu R. Tadikamalla. Amer. J. of Math. & Manag. Sci, V. 4, Nos. 3 & 4. American Sciences Pr, 1984, 209 pp, $49.75 (P). [ISBN: 0-935950-10-9] Special issue of this journal. Contains eight papers presenting state-of-the-art coverage, two input-related, two modeling-related, and four output-related. Note price. RSK

Statistics, T(17: 1, 2). Probability and Statistics, Volume I. Didier Dacunha-Castelle, Marie Duflo. Transl: David McHale. Springer-Verlag, 1986, vi + 362 pp, $32.50. [ISBN: 0-387-96067-8] Translation of the 1983 French edition. Sterile presentation of measure-theoretic probability and mathematical statistics. RSK

Statistics, S(17). Modern Concepts and Theorems of Mathematical Statistics. Edward B. Manoukian. Ser. in Stat. Springer-Verlag, 1986, xvi + 156 pp, $29.80. [ISBN: 0-387-96186-0] Collection of statistical facts, with references for further details, grouped into four chapters: basic definitions, concepts, results, and theorems; fundamental limit theorems; distributions; and some relations between distributions. RSK

Statistics, T(13: 1). Statistics: A Conceptual Approach. K. Laurence Weldon. Prentice-Hall, 1986, xix + 442 pp, $27.95. [ISBN: 0-13-845819-7] Written for non-science students, using a conversational style with a minimum of mathematical notation. Emphasizes descriptive statistics more than most, with minimal coverage of estimation and hypothesis testing. RSK

Statistics, T*(15-16: 1, 2). An Introduction to Mathematical Statistics and Its Applications, Second Edition. Richard J. Larsen, Morris L. Marx. Prentice-Hall, 1986, x + 630 pp, $36.95. [ISBN: 0-13-487174-X] Revision of the authors' 1981 text (TR, June-July 1981). Material has been added, examples have been updated, all exercises now occur at the ends of the appropriate sections, and other changes to make the text more readable have been made. It remains a well-written text emphasizing the interrelationship between probability theory, mathematical statistics, and data analysis. RSK

Statistics, P. Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series. K. Dzhaparidze. Transl: Samuel Kotz. Ser. in Stat. Springer-Verlag, 1986, vi + 324 pp, $49.50. [ISBN: 0-387-96141-0] Translation of the original Russian edition. Primarily concerned with testing hypotheses about the form of the spectral density function and estimating the unknown spectral parameters. RSK

Statistics, T(13-14). Statistical Analysis for Business and Economics, Third Edition. Donald L. Harnett, James L. Murphy. Addison-Wesley, 1985, xvii + 1011 pp. [ISBN: 0-201-10683-3] Revision of 1980 Second Edition adds chapters on analysis of variance and nonparametric statistics. Intended for a beginning course for business and economics students. Useful features include specified problems for computer use; special "study problems" with solutions; and accompanying workbook. KK

Statistics, T(16-18: 1), S, P, L. Cluster Dissection and Analysis: Theory, FORTRAN Programs, Examples. Helmuth Späth. Transl: Johannes Goldschmidt. Ser. in Comp. & Their Appl. Halsted Pr, 1985, 226 pp. [ISBN: 0-470-20129-0] Presupposes matrix algebra and FORTRAN. The theory of divisive cluster analysis, discussion of the algorithms needed, sample programs in FORTRAN, and examples. FLW

Statistics, P. Lecture Notes in Statistics-35: Linear Statistical Inference. Ed: T. Caliński, W. Klonecki. Springer-Verlag, 1985, vi + 318 pp, $26 (P). [ISBN: 0-387-96255-7] Proceedings of an International Statistical Conference on Linear Inference held in Poznań, Poland, in June 1984. Contains roughly half of the papers presented, representing its main features and results. RSK

Statistics, P*. Lecture Notes in Statistics-34: Metric Methods for Analyzing Partially Ranked Data. Douglas E. Critchlow. Springer-Verlag, 1985, x + 216 pp, $16.40 (P). [ISBN: 0-387-96288-3] Presents

new methods based on group theory for measuring the closeness of rankings when only a subset of the data is ranked, and illustrates their data-analytic applications on real data sets. Includes FORTRAN subroutines. RSK

Statistics, T(16-18: 1), S*, P*, L*. Robust Statistics: The Approach Based on Influence Functions. Frank R. Hampel, et al. Ser. in Prob. & Math. Stat. Wiley, 1986, xxi + 502 pp, $39.95. [ISBN: 0-471-82921-8] Presupposes calculus, linear algebra, and mathematical statistics. Robust estimation in theory and use from the point-of-view of influence functions. FLW

Statistics, T(13-14: 1). Elementary Statistics: In a World of Applications, Second Edition. Ramakant Khazanie. Scott Foresman, 1986, xii + 562 pp, $29.95. [ISBN: 0-673-16618-X] This revised edition (First Edition, TR, April 1979) includes coverage of stem-and-leaf diagrams, paired t-test, confidence intervals for prediction in regression, and p-values. Also more drill problems. FLW

Computer Literacy, P. The Digital Dictionary, Second Edition, Revised and Expanded. Ed: Robert E. Marotta. DEC, 1986, 659 pp, $25 (P). [ISBN: 0-932376-82-7] Digital here refers to Digital Equipment Corporation (DEC). This is a detailed guide to general and especially DEC-specific jargon. Definitions vary from "battery: noun, generic, a dc voltage source" to explanations of many DEC terms (and acronyms) in each of the hardware or operating system environments where DEC uses the terms. A survival kit for those who must read DEC manuals. JAS

Computer Literacy, P. AppleWorks: Applications. Lauren and Robert Flast. Osborne McGraw-Hill, 1986, x + 101 pp, $9.95 (P). [ISBN: 0-07-881236-4] Provides 25 business-related models to use with AppleWorks integrated software (word processor, spreadsheet, database). User must have working knowledge of AppleWorks, including use of formatting commands and clipboard communication. There is little explanation of how or why, simply a collection of "type in and use" applications. RD

Computer Literacy, P. AppleWorks: Tips & Traps. Dick Andersen, Janet McBeen, Janice M. Gessin. Osborne McGraw-Hill, 1986, xii + 291 pp, $16.95 (P). [ISBN: 0-07-881207-0] Provides hundreds of suggestions and cautions for using each component of the AppleWorks package (word processor, spreadsheet, database) and for communicating between components. Well written with many examples and illustrations, it is appropriate for experienced users as well as AppleWorks beginners. RD

Computer Literacy, T(13: 1). Computers and Applications: An Introduction to Data Processing. Daniel L. Slotnick, et al. DC Heath, 1986, xxxv + 755 pp, $29.95; $24.95 (P). [ISBN: 0-669-08675-4] Designed to be used as the textbook for the first course in a data processing curriculum. Well-organized and readable. AO

Computer Programming, S(13-15), L. The Students' FORTH. Glyn Emery. Comp. Sci. Texts. Blackwell Science, 1985, viii + 101 pp, (P). [ISBN: 0-632-01436-9] An introduction to the programming language FORTH for readers with prior programming experience in another language. Based on the FORTH-83 standard. Also includes implementation information. AO

Computer Programming, S(14-15). Advanced C: Food for the Educated Palate. Narain Gehani. Computer Science Pr, 1985, xiii + 313 pp, $19.95 (P). [ISBN: 0-88175-078-6] A concise introduction to the programming language C. Very similar to the author's C: An Advanced Introduction (TR, December 1985). More useful as a reference than as a textbook. AO

Computer Programming, S*(14-15), L. Modula-2, An Introduction. Daniel Thalmann. Springer-Verlag, 1985, xi + 292 pp, $19.50 (P). [ISBN: 0-387-13297-X] A concise, systematic introduction to Modula-2. Presents the complete language together with numerous examples. AO

Computer Programming, S(14-15), P, L*. A Guide to Modula-2. Kaare Christian. Texts & Mono. in Comp. Sci. Springer-Verlag, 1986, xix + 436 pp, $34. [ISBN: 0-387-96242-5] Designed to be both an introductory textbook and a reference. Organized by language features. Covers all aspects of the language in detail. AO

Computer Programming, T(13: 1). Standard Pascal: An Introduction to Structured Software Design. Victor J. Law. Wm C Brown, 1986, xvii + 558 pp, (P). [ISBN: 0-697-00080-X] Systematic approach; emphasis on programming methodology as well as language syntax; features presented following motivation for their use. Organized from conceptual view, e.g., sequence, selection, repetition control, modularity. Many complete programs, including specifications, algorithm design, structure charts. Some nice examples, exercises. RM

Computer Programming, T(13: 1, 2). Problem Solving with Pascal: Programming Methods, Algorithms, and Data Structures. James F. Peters III. Holt, Rinehart & Winston, 1986, xxii + 741 pp, $26.95 (P). [ISBN: 0-03-069848-0] An introductory textbook on Pascal programming following the ACM CS 1 and CS 2 course outlines. Appropriate for students with no prior computing experience. AO

Computer Programming, T(13: 1). PASCAL with Program Design. James F. Peters III. Holt, Rinehart & Winston, 1986, xxiv + 616 pp, $25.95 (P). [ISBN: 0-03-003282-2] Designed for a one-semester introductory course in Pascal programming that emphasizes a problem-solving approach to algorithm design. Coverage corresponds to the first half of the AP computer science course and the ACM CS 1 course. AO

Computer Programming, S*, P, L*. Programming Pearls. Jon Bentley. Addison-Wesley, 1986, viii + 195 pp, (P). [ISBN: 0-201-10331-1] Revisions of a baker's dozen of the author's popular column from Com-

munications of ACM emphasizing insight, creativity, and discipline in getting things right. Each column concludes with concise programming principles, problems (with hints and solutions in the back), and suggestions for further reading. A superb source of ideas and problems for courses in programming or data structures. LAS

Computer Programming, T*(13: 1, 2), L. Introduction to Pascal with Applications in Science and Engineering. Susan and Ellen Finger. DC Heath, 1986, xviii + 746 pp, $25.95 (P). [ISBN: 0-669-08609-6] For the first course in programming. Discusses many application topics in science and engineering in addition to (various versions of) Pascal itself. Many complete case studies. Debugging and style hints throughout. An attractive text containing a great deal of material. DFA

Computer Programming, S(14-15). APL, An Introduction. Howard A. Peelle. Holt, Rinehart & Winston, 1986, xvi + 459 pp, $27.95 (P). [ISBN: 0-03-004953-9] An informal introduction to a "generic" version of APL suitable for most systems. An earlier edition of the book was published by Hayden Books in 1978. AO

Computer Programming, T(13: 1, 2), L. PASCAL: An Introduction to the Art and Science of Programming. Walter J. Savitch. Benjamin/Cummings, 1984, xxiii + 550 pp, $20.95 (P). [ISBN: 0-8053-8370-0] Clearly written text for introductory programming course using either standard or UCSD Pascal. Covers procedures and parameters before loops and conditional statements. Chapters on recursion, numeric programming, data structures. Extensive discussion of programming techniques and style, testing, debugging. Assumes high school algebra. KS

Computer Programming, S(13-14), L. Macintosh, Graphics and Sound: Programming in Microsoft BASIC. David A. Kater. Osborne McGraw-Hill, 1986, x + 276 pp, $17.95 (P). [ISBN: 0-07-881177-5] Careful introduction to using Macintosh graphics and sound from Microsoft BASIC. Assumes a working knowledge of BASIC, but introduces from scratch the geometric and musical details necessary to understand the specialized Macintosh BASIC commands. LAS

Software Systems, T(16-17), L*. An Introduction to Database Systems, Volume I, Fourth Edition. C.J. Date. Addison-Wesley, 1986, xx + 639 pp. [ISBN: 0-201-14201-5] An extensively revised and updated edition of this widely used textbook (First Edition, TR, March 1976; Third Edition, TR, April 1982). Non-relational systems have been de-emphasized. Use of the IBM system DB2 has replaced System R and a chapter on INGRES has been added. AO

Software Systems, P, L*. LaTeX, A Document Preparation System. Leslie Lamport. Addison-Wesley, 1986, xiv + 242 pp, (P). [ISBN: 0-201-15790-X] LaTeX is a macro package for the typesetting system TeX that adds such commands as those that provide for automatic numbering of sections, theorems and references; for generation of indices and glossaries; for creation and cross-referencing of bibliographies; for simple pictures; and for preparation of multi-layer (colored) overhead transparencies. LaTeX is fully consistent with "raw" TeX, but only partially consistent with the common "plain" macro package used with TeX. LAS

Software Systems, P, L**.** TeX: The Program. Donald E. Knuth. Computers & Typesetting, B. Addison-Wesley, 1986, xv + 594 pp. [ISBN: 0-201-13437-3] The complete text of the TeX program, written in WEB, "a Pascal program that has been cut up into pieces and rearranged into an order that is easier for a human being to understand." The program consists of 1377 consecutively numbered single-topic sections, each of which contains commentary (in Knuth's typical wry style) introducing a chunk of formal Pascal code. Forward definitions in WEB make it possible to present the TeX program in a logical order and in the process "to explain the algorithms of TeX as clearly as possible." Handy mini-indices of identifiers appear on each facing pair of pages, giving section number references for every definition appearing on the pages. Second of five volumes in Knuth's TeX series. LAS

Software Systems, P, L**.** The METAFONTbook. Donald E. Knuth. Computers & Typesetting, C. Addison-Wesley, 1986, xi + 361 pp, [ISBN: 0-201-13445-4]; METAFONT: The Program. Computers & Typesetting, D. xv + 560 pp, [ISBN: 0-201-13438-1]; Computer Modern Typefaces. Computers & Typesetting, E. xv + 588 pp. [ISBN: 0-201-13446-2] Metafont is a program that creates typographical fonts--arrays of pixels for each symbol--for typesetting systems such as TeX using encoded Bezier curves and adjustable pen nib shapes to re-create calligraphic patterns traditionally drawn painstakenly by human artists and font designers. These three volumes, which complete Knuth's five-volume series on TeX and Metafont, contain a user's guide to Metafont, the complete program for Metafont written in WEB--a mixture of English and Pascal designed to make the program easy to understand--and a detailed case study giving Metafont programs and sample results for the 75 fonts that make up Knuth's Computer Modern typeface family--including Roman, slanted, typewriter, bold, italic sans serif, math italic, small caps, and math symbols in different sizes. Of interest not just to TeX users, but also as a well-documented example of a production-quality computer program and as a tool for artists and designers. LAS

Software Systems, S(16-18), P, L*. The INGRES Papers: Anatomy of a Relational Database System. Ed: Michael Stonebraker. Addison-Wesley, 1986, xi + 452 pp. [ISBN: 0-201-07185-1] A collection of papers summarizing the work of the INGRES project at the University of California, Berkeley, during the period 1973-1983. The papers discuss various issues in the design and implementation of a relational database system. AO

Computer Science, P, L*. A Guide to Expert Systems. Donald A. Waterman. Addison-Wesley, 1985, xviii + 419 pp. [ISBN: 0-201-08313-2] Written for readers without a background in artificial intel-

ligence or expert systems. Describes what expert systems are, how they work, and how they are used. AO

Computer Science, S(15-16), L. Ada in Practice. Christine N. Ausnit, et al. Books on Prof. Comput. Springer-Verlag, 1985, xv + 195 pp, $25 (P). [ISBN: 0-387-96182-8] A collection of guidelines and case studies illustrating proper use of the programming language Ada. Recommended for anyone teaching or using Ada. AO

Computer Science, T(14-15: 1), L. Structured Computer Organizations, Second Edition. Andrew S. Tanenbaum. Prentice-Hall, 1984, xiii + 465 pp, $39.95. [ISBN: 0-13-854489-1] An excellent introductory textbook for a first course on computer organization. Presents a hierarchical model of computer organization: the digital logic level, the microprogramming level, the conventional machine level, the operating system level, and the assembly language level. AO

Computer Science, P. Methodology of Window Management. Ed: F.R.A. Hopgood, et al. Eurographic Seminars. Springer-Verlag, 1986, xv + 250 pp, $34.50. [ISBN: 0-387-16116-3] Proceedings of a workshop held at the Rutherford Appleton Laboratory, April 29-May 1, 1985. Contains the invited presentations and reports from the working groups on issues related to the applications program interface, the user interface, and system architecture. AO

Computer Science, P. Lecture Notes in Computer Science-206: Foundations of Software Technology and Theoretical Computer Science. Ed: S.N. Maheshwari. Springer-Verlag, 1985, ix + 522 pp, $27.40 (P). [ISBN: 0-387-16042-6] Twenty-five of the papers presented at the fifth conference in this series that was held from December 16-18, 1985 in New Delhi. AO

Computer Science, P. Parallel Computers and Computations. Ed: J. van Leeuwen, J.K. Lenstra. CWI Syllabus, V. 9. Math Centrum, 1985, 184 pp, Dfl. 28.80 (P). [ISBN: 90-6196-297-8] Eight papers covering new developments (practical and theoretical) in parallel algorithms and architectures. Comparison of supercomputers (CRAY-1, CYBER 205), system programming, design and complexity of parallel algorithms, object-oriented distributed architectures, applications to numerical linear algebra, combinatorial optimization. RM

Computer Science, T(17-18: 1), P. Formal Techniques for Data Base Design. Antonio L. Furtado, Erich J. Neuhold. Springer-Verlag, 1986, viii + 114 pp, $24. [ISBN: 0-387-15601-1] A presentation of two different approaches representing current research in data base design: the application-oriented approach (emphasizing information, functions, and representation); and the semantic data approach emphasizing data structures. JAS

Computer Science, P. Lecture Notes in Computer Science-211: Complexity and Structure. Uwe Schöning. Springer-Verlag, 1986, 99 pp, $11.20 (P). [ISBN: 0-387-16079-5] Qualitative, structural complexity theory; emphasis on explanations for hardness of some problems rather than quantitative bounds. Complexity cores (instances uniformly hard); polynomially bounded circuit size complexity used to study the P = NP question; Berman-Hartmanis conjecture; reasons why some recursion theoretic principles (e.g., relativization) do not work as expected in complexity theory. RM

Computer Science, T(13-18: 1), S*, L*. Mastering C. Craig Bolon. Sybex, 1986, xii + 437 pp, $19.95 (P). [ISBN: 0-89588-326-0] This book falls half-way between being a high-quality programming book and being a programming language course treatment for C language. It is not a short-and-to-the-point tutorial; rather, it describes and explains C and how it works. It assumes some general computer programming experience and takes the time (space) to develop a number of concepts from the theory of programming languages in order to really explain how C works. However, it really teaches C programming, not language theory. Excellent index, no bibliography. JAS

Computer Science, P. Mathematical Logic and Programming Languages. Ed: C.A.R. Hoare, J.C. Shepherdson. Intern. Ser. in Comp. Sci. Prentice-Hall, 1985, 184 pp, $29.95. [ISBN: 0-13-561465-1] Proceedings of discussion meeting of Royal Society of London, held February 15-16, 1984. Ten papers on use of logic to design programming languages, formulate specifications for computer programs, and prove that programs meet their specifications. First published in Philosophical Transactions of the Royal Society, Series A, Volume 312. KS

Computer Science, P. Equational Logic as a Programming Language. Michael J. O'Donnell. Ser. in Found. of Comp. MIT Pr, 1985, 296 pp, $25. [ISBN: 0-262-15028-X] Description of theoretical foundations, design and implementation of logic programming language in which computations are produced from equational definitions. KS

Computer Science, P. Lecture Notes in Computer Science-200: Trace Theory and VLSI Design. Jan L.A. van de Snepscheut. Springer-Verlag, 1985, vi + 140 pp, $11.20 (P). [ISBN: 0-387-15988-6] This manuscript (in the literal sense) is the author's Ph.D. thesis. It is of unusually broad interest because of the fundamental nature of the ideas developed (concerning concurrent programming and chip design), and the unusual and refreshing style in which it presents a lot of mathematical methods for computer science. JAS

Applications, S*(13-16). UMAP Modules: Tools for Teaching 1985. COMAP, 1986, ix + 326 pp, (P). [ISBN: 0-912843-08-X] Eleven modules in applications of mathematics intended as supplements to a variety of standard mathematics courses. Topics range from continued fractions (in botany) and differential equations (applied to optical illusions) to the efficient design of computer circuits that add. Most are reprinted from the 1985 issues of The UMAP Journal. LAS

Applications (Artificial Intelligence), S(16), P. The Artificial Intelligence Experience: An Introduction. Susan J. Scown. Digital Pr, 1985, 183 pp, $15 (P). [ISBN: 0-932376-84-3] Non-technical introduction and overview of history, basics of current practice in artificial intelligence. Discussion of types of problems, languages, hardware available, systems in development or operation (e.g., DEC's XCON), standard techniques and paradigms, research efforts. RM

Applications (Artificial Intelligence), S(16). Artificial Intelligence: Applications to Logical Reasoning and Historical Research. Richard Ennals. Ser. in Comp. & Their Applic. Halsted Pr, 1985, 172 pp. [ISBN: 0-470-20181] Argument for the role of logic and logic programming (in PROLOG) in historical analysis, synergism between artificial intelligence and the humanities. Experiments with school children and PROLOG simulations of historical questions, information representation and retrieval; emphasis on PROLOG as tool for understanding explanation, interaction between historian and evidence. RM

Applications (Biology), P. Lecture Notes in Biomathematics-57: Mathematics in Biology and Medicine. Ed: V. Capasso, E. Grosso, S.L. Paveri-Fontana. Springer-Verlag, 1985, xviii + 524 pp, $36 (P). [ISBN: 0-387-15200-8] Proceedings of a conference held in Bari, Italy, July 18-22, 1983. Seventy papers on population genetics and ecology, epidemics, resource management, physiology and medicine, compartmental analysis, general mathematical methods. KS

Applications (Biology), P. Lecture Notes in Biomathematics-61: Resource Management. Ed: M. Mangel. Springer-Verlag, 1985, 138 pp, $11.40 (P). [ISBN: 0-387-15982-7] Presents the proceedings of the Second Rolf Yorque Workshop held in Ashland, Oregon, July 23-25, 1984. Addresses theoretical aspects of fishery production. Of limited interest to those outside the field. SM

Applications (Communication Theory), P. Maximum-Entropy and Bayesian Methods in Inverse Problems. Ed: C. Ray Smith, W.T. Grandy, Jr. Fund. Theories of Physics. D Reidel, 1985, ix + 492 pp, $44.50. [ISBN: 90-277-2074-6] Papers from two workshops held at the University of Wyoming from June 8-10, 1981, and from August 9-11, 1982. AO

Applications (Economics), P. Structural Sensitivity in Econometric Models. Edwin Kuh, John W. Neese, Peter Hollinger. Wiley, 1985, ix + 324 pp, $32.95. [ISBN: 0-471-81930-1] In the Wiley Series in Probability and Mathematical Statistics. Presents linear dynamic systems methods for interpreting complex macroeconomic models, and illustrates them using the Michigan Quarterly Econometric Model of the United States. RSK

Applications (Engineering), S(17), P. Design Optimization. Ed: John S. Gero. Notes & Reports in Math. in Sci. & Eng., V. 1. Academic Pr, 1985, xiv + 298 pp, $35. [ISBN: 0-12-280910-6] This volume contains nine different articles whose central theme is that design can be treated as a goal-seeking decision-making activity. The first five chapters deal with problems whose goals are expressed in a single objective while the last four chapters consider design problems where the goals are expressed using more than one objective. CEC

Applications (Engineering), T(17-18: 1). An Introduction to the Finite Element Methods with Applications to Nonlinear Problems. R.E. White. Wiley, 1985, x + 354 pp, $39.50. [ISBN: 0-471-80909-8] An elementary introduction to the finite element method written for graduate students in engineering. Emphasizes techniques for second-order linear and nonlinear partial differential equations. Presents techniques for incompressible viscous fluid flow problems and nonlinear heat transfer problems. AO

Applications (Engineering), P. A Geometric Investigation of Reach. James Urey Korein. MIT Pr, 1985, 208 pp, $30. [ISBN: 0-262-11104-7] New analytical solutions, algorithms, heuristics for the problem of human or robot arm movement. Polyhedral modelling of the reachable space, general problem of moving polyhedra through space, SurfsUP polyhedral modelling system discussed. 1984 ACM award dissertation. RM

Applications (Engineering), P. Computer-Aided Design in Magnetics. D.A. Lowther, P.P. Silvester. Springer-Verlag, 1986, viii + 324 pp, $45. [ISBN: 0-387-15756-5] Presents techniques for the effective use of CAD (computer-aided design) systems in the magnetic devices industry. Case studies are used to illustrate the basic principles. AO

Applications (Engineering), P. Topics in Boundary Element Research, Volume 2: Time-dependent and Vibration Problems. Ed: C.A. Brebbia. Springer-Verlag, 1984, xiv + 260 pp, $59. [ISBN: 0-387-13097-7] Boundary element methods are methods based on the integral equation formulation of initial-boundary value problems. This volume provides a survey of recent applications of these methods in elastodynamics. AO

Applications (Engineering), P. The Stationary Semiconductor Device Equations. Peter A. Markowich. Comput. Microelectronics. Springer-Verlag, 1986, ix + 193 pp, $45. [ISBN: 0-387-81892-8] A study of the static semiconductor device problem from the viewpoint of an applied mathematician. Provides an overview of mathematical results and techniques used in the design of device simulations. AO

Applications (Engineering), P. Probabilistic Models of Cumulative Damage. J.L. Bogdanoff, F. Kozin. Wiley, 1985, x + 341 pp, $53.50. [ISBN: 0-471-88180-5] Uses finite state discrete time imbedded Markov processes to model cumulative damage due to fatigue, fatigue crack growth, and wear, and illustrates their applicability with real data from a variety of sources. RSK

Applications (Engineering), P. Least Square Estimation with Applications to Digital Signal Processing. Arthur A. Giordano, Frank M. Hsu. Wiley, 1985, xxiii + 412 pp, $44.90. [ISBN: 0-471-87857-X] Part 1 presents least square error algorithms and the mathematical principles behind them. Part 2 describes applications involving equalization, spectral analysis, digital whitening, adaptive arrays, interchannel interference mitigation, and digital speech processing. RSK

Applications (Information Theory), T*(15-17: 1). Coding and Information Theory, Second Edition. Richard W. Hamming. Prentice-Hall, 1986, xii + 259 pp, $36.95. [ISBN: 0-13-139072-4] A concise, elementary introduction to error-detecting and error-correcting codes, to Shannon's theory of information, and (briefly) to algebraic coding theory. This Second Edition corrects and clarifies the First Edition, but refrains from adding to it: "it is better to master a little than to half-know a lot." Relatively few exercises. LAS

Applications (Medicine), P*. Clinical Trials: A Practical Approach. Stuart J. Pocock. Wiley, 1983, xii + 266 pp, $38.50. [ISBN: 0-471-90155-5] Straightforward, non-technical presentation of the basic principles of clinical trials. Emphasizes their scientific rationale, including the relevance of statistical methods. Also discusses ethical and organizational issues. Contains many interesting historical references and examples. RSK

Applications (Medicine),P*, L. Mathematical Methods in Medicine. Ed: D. Ingram, R.F. Bloch. Part I: Statistical and Analytical Techniques. Handbook of Applic. Math., Guidebook 3. Wiley, 1984, xix + 453 pp, $49.95 [ISBN: 0-471-90045-1]; Part II: Applications in Clinical Specialities. Handbook of Applic. Math., Guidebook 4. 1986, xxiii + 514 pp, $54.95. [ISBN: 0-471-90046-X] Collection of articles designed to describe how mathematics is used as a tool in medicine, illustrated with real examples. Treatment ranges from elementary to advanced, with numerous cross-references to the six core volumes in this Series which provide more detailed information about the concepts presented (TR, Series and Volume I, December 1980). Part I deals with general statistical and analytical techniques which find application in many clinical specialities, while Part II treats issues which tend to arise or have only been explored within one discipline. RSK

Applications (Meteorology), P. Lecture Notes in Statistics-30: Stochastic Models of Air Pollutant Concentration. Jan Grandell. Springer-Verlag, 1985, 110 pp, $11.40 (P). [ISBN: 0-387-96197-6] Summarizes and develops mathematical models for describing the variability of the concentration of pollutants in the atmosphere. RSK

Applications (Physics), S(13-15). Mathematical Physics. Robert Geroch. Lect. in Physics. U of Chicago Pr, 1985, vi + 351 pp, $15 (P); $30. [ISBN: 0-226-28862-5; 0-226-28861-7] Using category theory as a framework, this book introduces groups, vector spaces and linear algebra, the tensor product, Lie algebras, representation theory, topology, homology, measure theory and functional analysis. The book also includes exercises and brief discussions of special relativity, the algebra of observables in quantum theory, and dynamical systems. In order to accommodate its broad selection of topics, this book does compromise the depth and detail of its discussions. AM

Applications (Physics), P. Geometrodynamics Proceedings 1985. Ed: Agostino Pràstaro. World Scientific, 1985, vi + 466 pp, $41. [ISBN: 9971-978-63-6] Proceedings of the conference "Geometric Methods in Non-Linear Field Theory" held in Camigliatello Silano from January 15-19, 1985. AO

Applications (Physics), P. Lecture Notes in Physics-241: Numerical Simulation of Combustion Phenomena. Ed: R. Glowinski, B. Larrouturou, R. Temam. Springer-Verlag, 1985, ix + 404 pp, $26.80 (P). [ISBN: 0-387-16073-6] Proceedings of the symposium held at INRIA Sophia-Antipolis, France, May 21-24, 1985. JAS

Applications (Physics), T*(13-14: 1), S*, L. Quantum Mechanics in Simple Matrix Form. Thomas F. Jordan. Wiley, 1986, xi + 259 pp, $24.95 (P). [ISBN: 0-471-81751-1] A completely self-contained exposition of $QP-PQ = ih/2\pi$ and all that it implies using only school-level (precalculus) mathematics of complex numbers, matrix algebra, and expected value. Intended for gifted high school students or college students who are not majoring in physics, this echo of Max Born's matrix mechanics simplifies the calculus but not the insights of quantum mechanics. LAS

Reviewers

DFA: David F. Appleyard, Carleton; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; GF: Giovanna Fjelstad, St. Olaf; SG: Steven Galovich, Carleton; BH: Bruce Hanson, St. Olaf; PH: Paul Humke, St. Olaf; KK: Kenneth Kaminsky, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; AM: Alan Magnuson, St. Olaf; SM: Steve McKelvey, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

For the proof of (2), observe that

$$\begin{aligned}
&\frac{d}{dx}\binom{n}{k}F(x)^k(1-F(x))^{n-k}\\
&\quad=\left(\binom{n-1}{k-1}F(x)^{k-1}(1-F(x))^{n-k}-\binom{n-1}{k}F(x)^k(1-F(x))^{n-1-k}\right)nf(x)\\
&\quad=-\nabla\binom{n-1}{k}F(x)^k(1-F(x))^{n-1-k}nf(x).
\end{aligned}$$

Thus

$$\binom{n}{k}F(c)^k(1-F(c))^{n-k}=-\nabla\int_a^c\binom{n-1}{k}F(x)^k(1-F(x))^{n-1-k}nf(x)\,dx,$$

and by (5),

$$\begin{aligned}
\sum_{k=r}^{n}\binom{n}{k}F(c)^k(1-F(c))^{n-k}&=-\left[\int_a^c\binom{n-1}{k}F(x)^k(1-F(x))^{n-1-k}nf(x)\,dx\right]_{r-1}^{n}\\
&=\int_a^c\binom{n-1}{r-1}F(x)^{r-1}(1-F(x))^{n-r}nf(x)\,dx.
\end{aligned}$$

References

1. A. Chao, Another approach to incomplete integrals, The American Statistician, 36 (1982) 48.
2. W. Feller, An Introduction to Probability Theory and Its Applications, Vol 1, 3rd ed., Wiley, New York, 1968.

FINITE LINEAR GROUPS, THE COMMODORE 64, EULER AND SYLVESTER

DAVID BENSON, WALTER FEIT, AND ROGER HOWE

Department of Mathematics, Yale University, New Haven, CT 06520

For some examples in group theory, one of us was interested in knowing the number of conjugacy classes in $\mathrm{GL}_n(2)$, where $\mathrm{GL}_n(q)$ denotes the general linear group in dimension n over the finite field with q elements. A generating function was given in [3] for the number $p_n(q)$ of conjugacy classes in $\mathrm{GL}_n(q)$:

$$1+\sum_{n=1}^{\infty}p_n(q)x^n=\prod_{i=1}^{\infty}\frac{(1-x^i)}{(1-qx^i)}. \tag{1}$$

Since he had available a Commodore 64 minicomputer, the aforementioned one of us let it compute not simply $p_n(2)$, but the full polynomials $p_n(q)$, for $n < 40$. Here is a sample of the results.

$$\begin{aligned}
p_1(q)&=q-1,\\
p_2(q)&=q^2-1,\\
p_4(q)&=q^4-q,\\
p_8(q)&=q^8-q^3-q^2+q,\\
p_{16}(q)&=q^{16}-q^7-q^6-q^5+2q^3+q^2-q,\\
p_{32}(q)&=q^{32}-q^{15}-q^{14}-q^{13}-q^{12}-q^{11}-q^{10}+q^9\\
&\quad+2q^8+4q^7+3q^6-4q^4-3q^3+q^2+q.
\end{aligned} \tag{2}$$

Evidently the Commodore 64 had uncovered something interesting. The polynomials $p_n(q)$ are very atypical, in that the coefficient of q^k vanishes for $n > k > (n-1)/2$, and equals -1 for $(n-1)/2 \geqslant k > (n-3)/3$. Particularly interesting from the group-theoretical viewpoint is the fact that q^n is an astoundingly good approximation to $p_n(q)$. Evidently a tremendous amount of cancellation occurs if $p_n(q)$ is computed using formula (1).

Intrigued by this behavior, another of us studied the Commodore's initial output, which was for $n < 20$, and conjectured the following formula:

$$(3)\quad \prod_{i=1}^{\infty} \frac{1-x^i}{1-qx^i} = 1 + \sum_{m \geqslant 0} q^m \left(\sum_{\substack{r \geqslant 0 \\ (m,r) \neq (0,0)}} (-1)^r x^{(r+1)m + r(3r-1)/2} \frac{(1-x^{m+2r})}{(1-x^{m+r})} \binom{m+r}{r}(x) \right),$$

where

$$(4)\qquad \binom{m+r}{r}(x) = \frac{(1-x^{m+r})(1-x^{m+r-1}) \cdots (1-x^{m+1})}{(1-x^r)(1-x^{r-1}) \cdots (1-x)}.$$

The functions $\binom{m+r}{r}(x)$ are easily checked to satisfy the recursions

$$(5)\qquad \binom{m+r}{r}(x) = \binom{m+r-1}{r}(x) + x^m \binom{m+r-1}{r-1}(x).$$

(See [6], p. 175.) Using (5) one sees by induction that $\binom{m+r}{r}(x)$ is a polynomial of degree mr with nonnegative coefficients. Hence there is relatively little cancellation of coefficients in the right-hand side of (3), so that it provides a much faster way of computing the polynomial $p_n(q)$ than does the deceptively compact right-hand side of (1). For example the next polynomial in the series (2) was computed by hand from (3) to be

$$(6)\quad p_{64}(q) = q^{64} - \sum_{i=21}^{31} q^i + 2q^{19} + 3q^{18} + 5q^{17} + 6q^{16} + 8q^{15} + 7q^{14} + 4q^{13} - 2q^{12} - 10q^{11}$$
$$-17q^{10} - 17q^9 - 6q^8 + 10q^7 + 20q^6 + 9q^5 - 6q^4 - 7q^3 + q.$$

Further computation by the Commodore 64 verified formula (3) up to $n = 39$, so we attempted to prove (3). We briefly sketch our derivation with comments. We thank George E. Andrews for providing us with the historical remarks below.

We begin with two famous identities of Euler.

$$(7)\qquad \prod_{i=1}^{\infty}(1-x^i) = 1 + \sum_{r=1}^{\infty}(-1)^r\left(x^{r(3r+1)/2} + x^{r(3r-1)/2}\right) = \sum_{r=-\infty}^{\infty}(-1)^r x^{r(3r-1)/2}$$

$$(8)\qquad \sum_{i=1}^{\infty} \frac{1}{1-qx^i} = \sum_{m=0}^{\infty} q^m \frac{x^m}{(1-x^m)(1-x^{m-1}) \cdots (1-x)}.$$

See [5] p. 284 and [1] p. 19. Note that formula (7) results from (3) by specializing q to be zero. From (8) we can write

$$(9)\qquad \prod_{i=1}^{\infty} \frac{1-x^i}{1-qx^i} = \sum_{m=0}^{\infty} q^m x^m \prod_{j=m+1}^{\infty}(1-x^j).$$

If the product in the right-hand side of (9) is expanded one sees the coefficient of q^m is

$$x^m - x^{2m+1} - x^{2m+2} \cdots - x^{3m} + \text{higher terms}.$$

This explains the most obvious features of the coefficients of $p_n(q)$. However the right-hand side of (9) still involves much more cancellation than the right-hand side of (3). We can get (3) from (9) if the identity

$$(10)\qquad \prod_{j=m+1}^{\infty}(1-x^j) = \sum_{r=0}^{\infty}(-1)^r x^{mr + r(3r-1)/2} \frac{(1-x^{m+2r})}{(1-x^{m+r})} \binom{m+r}{r}(x)$$

holds. (In (10) if $m = r = 0$, we agree that $(1 - x^0)/(1 - x^0) = 1$.) Or multiplying both sides of equation (10) by $\prod_{j=1}^{m}(1 - x^j)$ we get the equivalent relation

$$\prod_{j=1}^{\infty}(1 - x^j) = \sum_{r=0}^{\infty}(-1)^r A_m(r), \tag{11}$$

where the $A_m(r)$ are polynomials given by

$$A_m(r) = x^{mr + r(3r-1)/2}\frac{(1 - x^{m+2r})}{(1 - x^{m+r})}\prod_{l=1}^{m}(1 - x^{r+l}).$$

(Note that for $m = 0$, the product $(1 - x^{m+r}) \cdots (1 - x^{r+1})$ equals 1, while for $m \geqslant 1$, the terms $(1 - x^{m+r})$ cancel.)

As we have remarked, equation (11) for $m = 0$ is just Euler's identity (7). It turns out that the generalization (11) of (7) is a special case of a result of Sylvester [7] and was put into the more elegant form

$$\prod_{j=1}^{\infty}(1 - x^j) = \sum_{r=-\infty}^{\infty}(-1)^r x^{mr + r(3r-1)/2}(1 - x^{r+m-1}) \cdots (1 - x^{r+1}) \tag{12}$$

(corresponding to the third expression in (7)) relatively recently by I. J. Good [4]. Modulo notation, Sylvester's generalization of (11) can be found in [**1**]. p. 140.

For the benefit of the reader who, with us, is not familiar with Sylvester's identity, we indicate the proof we devised of (11). By induction on m, beginning with $m = 0$ (Euler's identity (7)), it is enough to show that the right-hand sides of (11) for m and $m + 1$ are equal. In other words, we want to show

$$\sum_{r=0}^{\infty}(-1)^r(A_{m+1}(r) - A_m(r)) = 0.$$

This will follow by formal manipulation if we can find polynomials $B_m(r)$ such that $B_m(0) = 0$ and

$$A_{m+1}(r) - A_m(r) = B_m(r+1) + B_m(r). \tag{13}$$

But, indeed, if we set

$$B_m(r) = -x^{mr + r(3r-1)/2}\prod_{j=r}^{m+r-1}(1 - x^j)$$

for $m \geqslant 0$, $r \geqslant 0$, $mr \neq 0$, and $B_0(0) = 0$, then the verification of (13) is straightforward. The reader should note the similarity between the $B_m(r)$ and the terms of the sum in (12). This argument for identity (11) turns out to be essentially like one given by Sylvester's friend, Arthur Cayley [**2**], Vol. 12, pp. 217–219.

Thus these classical identities, long cultivated for their own sake, find an application in group theory. Although formulas analogous to (1) are known for the other classical groups over finite fields, none of these other formulas seem to give rise to intricate cancellations of the sort yielding expressions as in (2).

References

1. G. E. Andrews, The Theory of Partitions, Addison-Wesley, Reading, MA, 1976.

2. A. Cayley, Collected Papers.

3. W. Feit and N. J. Fine, Pairs of commuting matrices over a finite field, Duke Math. J., 27 (1960) 91–94.

4. I. J. Good, A generalization of the Bernoulli-Euler partition formula, Scripta Math., 28 (1968) 319–320.

5. G. M. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, 4th Ed., Oxford University Press, Oxford, England, 1959.

6. T. Nagel, Introduction to Number Theory, Wiley, New York, 1951.

7. J. J. Sylvester, Collected Papers, 4 (1883) 91, 93–94.

A CLASS OF RINGS WHICH ARE VERY NEARLY ASSOCIATIVE

ERWIN KLEINFELD
Department of Mathematics, The University of Iowa, Iowa City, IA 52242

In nonassociative ring theory the word ring is used to denote a system which satisfies all the usual axioms except the associative rule of multiplication. The very first example of such a ring seems to have arisen in the middle of the last century in the form of the Cayley numbers (octonians) [**1**]. The Cayley numbers have the property that

$$(x, y, y) = 0 = (y, y, x),$$

where the associator (x, y, z) is defined by

$$(x, y, z) = (xy)z - x(yz)$$

for x, y, and z elements of a ring. Circa 1930, Artin and Zorn defined an alternative ring as a ring satisfying

$$(x, y, y) = 0 = (y, y, x).$$

One learns how a piece of machinery works by taking it apart and putting it together again. In similar fashion an algebraist might test the definition of alternative ring by trying to classify all the basic examples. In ring theory, both associative and nonassociative, this has come to mean finding all simple ones. A ring R is defined to be *simple* if it has no ideals other than 0 and R and $R^2 \neq 0$. A ring is called *semiprime* if it has no ideal $I \neq 0$ such that $I^2 = 0$. Clearly simple implies semiprime but the converse is false.

While Zorn was classifying simple alternative rings with finiteness conditions, the same identities that had been used to define an alternative ring arose in some work Moufang was doing on projective planes [**4**]. By the fifties enough was known about alternative rings to provide algebraic proofs of some interesting results in projective geometry [2], [3].

Jordan rings are defined as commutative rings which in addition satisfy the identity $(x^2, y, x) = 0$. The original justification for studying such rings was that they were going to be helpful in explaining quantum mechanics. However, before the ink was dry on the first paper, physicists had already lost interest in Jordan rings. Nevertheless Jordan and alternative rings are still very much alive today. Nonassociative rings without additional identities have never been in the limelight, perhaps because there are so few interesting theorems known about them. Even the finite division rings are far from classified. Quite recently some people in nuclear and Hadron physics have become interested in rings that are called Lie-admissible. These are rings which satisfy the identity

$$(x, y, z) + (y, z, x) + (z, x, y) - (x, z, y) - (z, y, x) - (y, x, z) = 0.$$

In an associative ring, all associators are zero. In studying nonassociative rings it is natural to ask what happens if all associators have certain special properties. Thedy has studied rings in which it is assumed that associators commute with all elements [**8**]. In this spirit we shall consider a class of rings in which all associators are in the nucleus. The nucleus of a ring R consists of all elements n in R such that

$$(n, r, s) = 0 = (r, s, n) = (s, n, r), \quad \text{for all } r, s \text{ in } R. \tag{1}$$

We shall prove the following assertion.

THEOREM. *Let all associators of R lie in its nucleus. Suppose that* $2x = 0$ *implies* $x = 0$ *and that R is semiprime. Then R must be associative.*

The proof we are about to present is easy to comprehend and not unduly computational, yet imparts the general flavor of those other papers. A sampling of some of these papers appears

among the references.

Now for the proof of the theorem.

Proof. We will need to use an identity valid in all rings, the so-called Teichmüller identity,

$$(wx, y, z) - (w, xy, z) + (w, x, yz) = w(x, y, z) + (w, x, y)z. \tag{2}$$

To verify this identity, we expand the left-hand side using the definition of the associator. We now take turns letting one of the four elements in (2) be in the nucleus. Thus

$$(nx, y, z) = n(x, y, z), \tag{3}$$

$$(wn, y, z) = (w, ny, z), \tag{4}$$

$$(w, xn, z) = (w, x, nz), \tag{5}$$

$$(w, x, yn) = (w, x, y)n, \tag{6}$$

for n any element in N and the rest of the elements arbitrary in R. The reader may wish to verify that N is a subring at this point. To do so one should verify the linearity of the associator and then assume x is in N in (3), y is in N in (4) and z is in N in (5).

All rings R have an ideal A, called the associator ideal. It is defined as the smallest ideal which contains all associators. It actually consists of all finite sums of associators and right multiples of associators. To see that the elements of this form constitute an ideal we observe that

$$(w, x, y)z \cdot s = ((w, x, y), z, s) + (w, x, y) \cdot zs,$$

is again of this form, while

$$z(w, x, y) = (zw, x, y) - (z, wx, y) + (z, w, xy) - (z, w, x)y,$$

using (2), is also of the desired form. Noting that

$$s \cdot (w, x, y)z = -(s, (w, x, y), z) + s(w, x, y) \cdot z$$

and using the previous two identities, we see that this is again of the desired form. The details are left to the reader.

The associator ideal is never zero, except when R is associative. Let $q = (a, b, c)(x, y, z)$, where a, b, c, x, y, z are arbitrary elements of R. Using (a, b, c) as the nuclear element n in (3), we see that $q = ((a, b, c)x, y, z)$. At this point (2) implies

$$(a, b, c)x = -a(b, c, x) + (ab, c, x) - (a, bc, x) + (a, b, cx).$$

Inserting this equation on the left-hand side of the associator with y and z, as well as using the hypothesis that associators are in the nucleus, we see that

$$q = -(a(b, c, x), y, z).$$

Now identity (4) with $n = (b, c, x)$ shows that $q = -(a, (b, c, x)y, z)$. Then (2) again implies $q = (a, b(c, x, y), z)$. Then (5) gives $q = (a, b, (c, x, y)z)$. At this point (2) implies $q = -(a, b, c(x, y, z))$. Consequently (6) may be used to show $q = -(a, b, c)(x, y, z)$.

Looking back at how we defined q to begin with, we have now verified that $q = -q$. Since by hypothesis, there are no elements of additive order two, it follows that $q = 0$. Now the stage is set for establishing that the associator ideal A of R squares to zero. Multiplying equation (2) on the right by (a, b, c), using the fact that all associators are in the nucleus by hypothesis and that we have just proved that every product of two associators is zero, we see that

$$(w, x, y)z(a, b, c) = 0. \tag{7}$$

Then $[(w, x, y)z][(a, b, c)d] = [(w, x, y)z(a, b, c)]d$, since (a, b, c) lies in the nucleus. But then use of (7) gives us

$$[(w, x, y)z][(a, b, c)d] = 0.$$

This in essence is all that is required to show $A^2 = 0$. Since R is semiprime we therefore obtain $A = 0$, which means R must be associative. This completes the proof of the theorem.

How can one construct an example of a ring all of whose associators lie in its nucleus? One starts with the free nonassociative ring on k generators, $k \geq 1$. Then set all words of length five or more equal to zero. Since the hypothesis of all associators in the nucleus is a trio of identities, each involving words of length five or more, we have produced such a ring. Incidentally in this ring $(x, x, x) \neq 0$, for each generator x.

References

1. A. A. Albert (Editor), M. A. A. Studies in Mathematics, Studies in Modern Algebra, Vol. 2, Mathematical Association of America, Washington, DC, 1963.
2. R. H. Bruck and Erwin Kleinfeld, The structure of alternative division rings, Proc. Amer. Math. Soc., 2 (1951) 878–890.
3. R. H. Bruck, 4th Slaught Memorial Paper, Contributions to Geometry, Recent advances in Euclidean plane geometry, this MONTHLY, 62 (1955) 2–17.
4. Ruth Moufang, Alternativekörper und der satz vom Vollständigen vierseit (D_a), Abh. Math. Sem. Univ. Hamburg, 9 (1933) 207–222.
5. R. D. Schafer, An Introduction to Nonassociative Algebras, Academic Press, New York, 1966.
6. A. K. Suschkewitsch, On a generalization of the associative law, Trans. Amer. Math. Soc., 31 (1929) 204–214.
7. Armin Thedy, On rings with commutators in the nuclei, Math. Z., 119 (1971) 213–218.
8. ______, On rings satisfying $[(a, b, c), d] = 0$, Proc. Amer. Math. Soc., 29 (1971) 250–254.
9. ______, Ringe mit $x(yz) = (yx)z$, Math. Z., 99 (1967) 400–404.
10. K. A. Zhevlakov, A. M. Slinko, I. P. Shestakov, and A. I. Shirshov, Rings That Are Nearly Associative (translated by Harry F. Smith from Russian), Academic Press, New York, 1982.

π, e, AND OTHER IRRATIONAL NUMBERS

ALAN E. PARKS

Department of Mathematics, Lawrence University, Appleton, WI 54912

In [1] Niven gave a clever, short proof that π is irrational. We would like to show how his proof can be generalized to prove substantially more:

(a) If $0 < |r| \leqslant \pi$ and if $\cos(r)$ and $\sin(r)$ are rational, then r is irrational.
(b) If r is positive and rational, $r \neq 1$, then $\ln(r)$ is irrational.

For example, (a) shows that π is irrational. If $a^2 + b^2 = c^2$ for rational numbers a, b, c, with $bc \neq 0$, then (a) shows that $\arccos(a/c)$ is irrational. Taking the contrapositive in (b) with $r = e$, we see that e is irrational.

Of course, by a famous theorem of Lindemann, all the numbers in (a) and (b) are not just irrational, but transcendental. The novelty of our argument is not so much the conclusion, but that our proof is elementary and can be effectively presented to students of calculus, for we require nothing beyond integration by parts and knowledge of the limit $\lim_{k\to\infty} M^k/k!$. Then again, at the beginning level, the fact that certain real numbers which occur naturally are irrational is interesting enough to present for its own sake.

We will apply the following theorem to prove (a) and (b). Its proof is a generalization of Niven's argument in [1].

THEOREM. *Let c be a positive real number and let $f(x)$ be a continuous function on $[0, c]$, positive on $(0, c)$. Suppose there are (antiderivatives) $f_1(x), f_2(x), \ldots$ with $f_1'(x) = f(x)$ and with $f_k'(x) = f_{k-1}(x)$ for all $k \geqslant 2$, and such that $f_k(0), f_k(c)$ are integers for all $k \geqslant 1$. Then c is irrational.*

Proof. Let P be the set of all polynomials $g(x)$ with real coefficients such that $g(0), g(c), g'(0), g'(c), \ldots, g^{(k)}(0), g^{(k)}(c), \ldots$ are all integers.

CLAIM 1. *If $g(x)$ is in P, then $\int_0^c f(x)g(x)\,dx$ is an integer.*

Proof. Successive integrations by parts give

$$\int_0^c f(x)g(x)\,dx = \left[f_1 \cdot g - f_2 \cdot g' + f_3 \cdot g'' - \cdots + (-1)^d f_{d+1} \cdot g^{(d)}\right]_0^c,$$

where d is the degree of $g(x)$. This proves the claim.

We will also need the following easy fact.

(1) If $g(x)$ and $h(x)$ are in P, then so is $g(x)h(x)$.

Now assume that c is rational, and write $c = m/n$, where m, n are positive integers. Then one verifies:

(2) $m - 2nx$ is in P.

Let $g_k(x) = x^k(m - nx)^k/k!$ for $k = 0, 1, 2, \ldots$.

CLAIM 2. *$g_k(x)$ is in P for all k.*

Proof. Induction on k: $g_0(x) = 1$ is an element of P. For $k \geqslant 1$,

$$g_k'(x) = g_{k-1}(x)(m - 2nx).$$

By induction, g_{k-1} is in P, by (2) $m - 2nx$ is in P, and thus by (1) g_k' is in P. Since also $g_k(0)$ and $g_k(c)$ are 0, we have that g_k is in P.

Observe that $g_k(x) > 0$ on $(0, c)$, a property shared by $f(x)$, so that $\int_0^c f(x)g_k(x)\,dx > 0$. By Claim 1, the integral is also an integer; therefore

(3) $\int_0^c f(x)g_k(x)\,dx \geqslant 1$ for all k.

Let M be the maximum for $x(m - nx)$ on $[0, c]$, and L that for $f(x)$, then

$$\int_0^c f(x)g_k(x)\,dx \leqslant \int_0^c L \cdot \frac{M^k}{k!}\,dx = c \cdot L \cdot \frac{M^k}{k!}.$$

But $\lim_{k\to\infty} M^k/k! = 0$, contradicting (3). We are forced to conclude that c is irrational.

To prove the statement (a) mentioned at the beginning, observe that if $\cos(r)$ and $\sin(r)$ are rational, so are $\cos(|r|)$ and $\sin(|r|)$, and thus we can find a positive integer n such that $n \cdot \sin(|r|)$ and $n \cdot \cos(|r|)$ are integers. Apply the theorem, with $c = |r|$ and $f(x) = n \cdot \sin(x)$, to conclude that $|r|$ is irrational, hence that r is irrational.

To prove (b), observe that $r > 1$ without loss of generality, so that $\ln(r) > 0$. Write $r = m/n$ for some positive integers m, n, and apply the theorem with $c = \ln(r)$ and $f(x) = n \cdot e^x$.

Reference

1. I. Niven, A simple proof of the irrationality of π, Bull. Amer. Math. Soc., 53 (1947) 509.

ANSWER TO PHOTO ON PAGE 715

Kazimir Kuratowski.

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP
For instructions about submitting material for publication in this department see the inside front cover.

SUCCESSFUL REMEDIAL MATHEMATICS PROGRAMS: WHY THEY WORK

DENA PATTERSON
Group in Science and Mathematics Education, University of California, Berkeley, CA 94720

TOM SALLEE
Department of Mathematics, University of California, Davis, CA 95616

In the past decade, the need for college remedial courses in mathematics has grown to unprecedented levels throughout the nation. Several states are attempting to institute various corrective measures at the high school level, but all evidence indicates that the problem of mathematical remediation will continue to plague colleges and universities for the foreseeable future.

It was in response to this problem that the University of California and the California State University jointly developed a reliable diagnostic test to assist students in choosing the appropriate mathematics course upon their entry to college. Each year approximately 40% of those tested are advised to repeat the material from one or more of their high school mathematics classes and, as a consequence, large numbers of remedial mathematics courses have sprung into existence. Unfortunately, most have been developed without adequate models or extensive support, and as a result, they are unable to do the task they were designed for.

To address this issue, a work group was formed under the auspices of the UC/CSU Joint Projects Task Force and was charged with investigating promising remedial programs in mathematics nationwide and proposing model programs that might be adopted in two California systems. Ultimately this investigation involved a literature review, a telephone survey of over a hundred institutions and site visits to 20 campuses. The directors of the most promising programs were invited to share their expertise with representatives from the two California university systems at a conference in April, 1984. The purpose of this article is to summarize the findings of the study, by reporting the common elements of successful programs, providing an overview of some of the issues of importance in remediation and categorizing the types of programs that have been developed in response to these issues.

Keys to Successful Programs. While the style and specific details of the successful programs vary considerably, there is agreement across the country regarding the extent of the problem and certain aspects of its potential solutions. One message emerged clearly during the study: even the best remedial program will wither or die without both strong administrative support and a competent, dedicated person at its helm.

A solid commitment to remediation as a relatively permanent necessity in the institution is the first, and maybe the most important, component required to establish an effective remedial program. This commitment must include a reasonable and permanent funding base, administrative backing from high levels and an appropriate physical environment. The lack of long term commitments to program funding and to the people who develop and run the remedial courses have been major reasons for both the lack of development of effective remedial programs and the early demise of many successful ones.

Developing and coordinating effective remedial programs requires a great deal of time and energy from a dedicated, creative faculty member. Such a person must be found and convinced to undertake the enterprise. Unfortunately, then, in many institutions the effort needed to design and run a good remedial program goes unrewarded by promotions and raises, and the faculty involved become unwanted step-children within their own departments. Not too surprisingly, most faculty

would rather not become involved in such a professionally fruitless endeavor, and many abandon their efforts and return to the better-rewarded, more traditional enterprises as soon as possible. To produce high quality, effective, remedial programs, those faculty who are involved in their development and coordination must be granted the promotions, raises and status which would normally accrue to them as mathematics faculty. Otherwise, it is unlikely that these programs will ever show continued success.*

Beyond the support of an administration committed to an on-going quality program of remediation, and a dedicated, hard-working faculty member to design and coordinate the program, other factors are of importance to a successful remedial program as well. Those which appear to be most significant are the careful sequencing of the courses, accurate placement of the students, efficient use of student tutors and TA's, and computer management of the bookkeeping.

Remedial programs need a well-integrated sequence of courses leading to clearly defined entry courses in the college mathematics sequence. In particular, there must not be a sizable gap in the level of sophistication between the end of one course and the beginning of the next. While strong students may be able to make up the difference on their own, weak students cannot. In order to direct students to the appropriate courses, programs need well-developed placement tests, carefully geared to their own sequence of courses. With a carefully designed sequence of courses, and placement exams that coordinate well with this sequence of courses, the placement needs of the students can be determined and met. This proper placement ensures that students pass through the program in the shortest possible time.

Every program encountered in the study relied on student assistants to some degree. In fact, undergraduate tutors and teaching assistants, carefully chosen, trained and supervised, formed the core of the most successful programs. It was consistently reported that students in disciplines other than mathematics make the best tutors and TA's, because they tend to be more sympathetic and patient with the remedial students' problems than most mathematics students. Besides needing a reasonably thorough understanding of mathematics, a good tutor or TA for a remedial program must have strong "people skills" such as empathy, patience and a sense of humor. In addition, tutors must be willing to critically review what they do in the process of tutoring and accept feedback and direction. Videotaped presentations given by the applicants and critiqued by the experienced tutors and TA's have proven to be an effective way of identifying these traits during the tutor selection process. In this way much can be learned about the applicant's approach to teaching and about his or her ability to accept criticism and improve through it.

While a few tutors and TA's may flourish in their teaching unaided, most need, and all may benefit from, a strong and effective program of training and supervision. Tutor and TA training programs range from a few hours of introduction to the university's remedial program to semester-long classes involving readings, role playing and class projects. Supervision is usually done entirely by the director of the program or the instructor in charge of the class. A clear structure for the supervision is best; otherwise, it often degenerates into a somewhat haphazard matter of giving occasional direction and dealing with emergencies as they arise. The most effective supervision program found in the study took the form of mentor relationships set up by the program director between the experienced TA's and the beginning TA's. (The mentor program and videotape selection process were both found in Harvard's Math AR program under the direction of Deborah Hughes-Hallett.)

Effective use of computer assisted instruction (CAI) is still in the developmental state in most institutions. But any remedial mathematics program serving more than a few hundred students will find computer managed instruction (CMI) a necessity. The most sophisticated of these programs keep all class records, generate examinations for individual students, grade the students' responses, and monitor their overall progress, signaling those that are in particular need of

*As John Gardiner put it in *Excellence*: "The society which scorns excellence in plumbing because plumbing is a humble activity and tolerates shoddiness in philosophy because it is an exalted activity will have neither good plumbing nor good philosophy. Neither its pipes nor its theories will hold water."

intervention. There is even one CMI program which periodically gives each student's projected course grade based on the performance of students in past years (Northern Arizona University).

A number of other issues consistently surfaced in the study and call for further investigation. Among these are: how to evaluate the effectiveness of a program in terms of the success of its students in future math courses; how to determine and compare the cost-effectiveness of the different programs; and how best to deal with the affective problems of remedial students, such as resentment of the course requirement, anxiety and procrastination.

Types of Programs. While there were wide individual variations, the programs investigated clustered naturally into three general types based upon the program's primary method of imparting information to the student, the amount and type of student-teacher interaction typical in the program, and the amount of structure that the program imposed on the student on a day-to-day basis. We term these program types lecture classes, independent learning courses, and individualized classes without lectures.

In lecture classes the instructor is viewed as the primary source of information, with the text and other sources serving supplementary roles. Student/teacher interaction varies depending on the initiative of the student and the instructor. Class meetings, homework assignments and test schedules pace the course too fast for some and too slow for others, but provide the high degree of structure often needed by remedial students. It is usually hoped that greater maturity and greater incentive on the part of the student will combine to produce success in the college remedial mathematics class.

In the independent learning course, the students are generally given access to several means of learning the material and the responsibility for acquiring the necessary information is left to the students' initiative. Student/teacher interaction may be minimal or non-existent. The students progress at their own pace, and generally have a minimum amount of day-to-day structure imposed by the course. As a result, procrastination is frequently a major problem for students in these programs. Maturity and initiative are required of the student, and may be developed in the process of completing the course. Many self-paced programs find that large percentages of students drop the course, while those that remain take several semesters to complete it. On the brighter side, those students who do complete the course usually exhibit a thorough knowledge of the material.

A third approach to remedial mathematics courses, individualized classes without lectures, requires a prescribed number of hours per week in a class or a learning center. In a group setting individuals study worktexts and supplementary sources, such as audio-visual and computer materials while an instructor and tutors are available for help with the material as needed. This arrangement provides high degrees of both structure and of interaction between the individual and the instructor for those students who need these.

Model Programs. As a result of the telephone survey and site visits, five programs were identified which exemplify the most effective features of each of these program types. These model programs are identified and described below.

Lectures Revisited. Students in remedial math courses often have trouble dealing with formalization and abstraction, a difficulty which probably impeded their learning the material in high school. While some courses focus on the formalizations and hope that one more exposure will help these ideas make sense to the students, most do not emphasize abstractions, but concentrate on skills. Unfortunately rote skills generally do not mean much more to the remedial student than the abstractions do. Neither of these approaches takes into account that factor which is usually the main problem for these students: namely, their level of cognitive development in mathematics.

The most successful lecture-type programs that we found were those which deal honestly with the ability of the remedial math student to handle abstraction. These courses present material in concrete ways that are accessible to the students. Once basic concepts are developed in a familiar context, the student is then led into more formal and abstract thought.

For example, at Ohio State University, the course "Transition to College Math" developed by

Joan Leitzel, begins with a highly numerical approach, using calculators as essential equipment. It progresses to a visual, graphical approach and then moves to equations and functions. According to the designer of another successful program, Gabriella Wepner at Ramapo College, "instruction proceeds from the concrete or intuitive, to the abstract, emphasizing process rather than specific rules. ... Students are continually asked: to look for patterns in examples; to explain any relationships they see; to test if the proposed relationship is maintained in other examples, if not to propose alternatives; and finally to make generalizations based on their observations. ... Since the abstract formalization is the result of the students' own experience, it is more easily assimilated. ... Consequently, students demonstrate significant gains at the end of instruction" [**2**]. This course has an impressive 90% success rate.

Independent Learning. The Math Module System at Colorado State University is a very carefully developed, cost-effective program based on the PSI-Keller Plan of instruction. The program consists of seven modules of pre-calculus material, from which the student may choose those modules that are appropriate to his or her goals. There are texts, lectures, video-tapes and tutors available for the students to learn from at their own discretion.

In order to receive credit for a module, the student must pass carefully developed multiple-choice tests that require a solid understanding of the material for successful completion. There are more than fifty tests for each unit, so that students may retake tests as often as necessary. The tests thus become an important learning tool for the students. Since there is no day-to-day structure imposed on the student, procrastination is reported to be a major problem for the less mature and less well-prepared student.

Individualized Classes. Approaches to remedial mathematics courses have been developed which impose a good deal of structure on the student and allow a high degree of student/teacher interaction, while leaving the responsibility for acquisition of information with the student.

At American River College in Sacramento, remedial math classes meet in the Mathematics Learning Center. Students use workbooks and audiotapes under the supervision of their instructor and tutors. Students can follow either of two test deadline schedules, thus completing the course either in one semester or two. This format provides the structure and student-teacher interaction that these students so often need, while allowing for some elements of self-pacing.

Northern Arizona University has a similar program, in which students work out of workbooks under the supervision of their instructor and tutors. Class attendance is mandatory. The test deadlines must be met, although a student may complete the course early. This format provides the structure found in the traditional lecture classes and allows greater individual student/teacher interaction.

Summary. In summary, an overwhelming number of students with vastly different needs and abilities are funneled into many different kinds of remedial programs each year. Many types of programs can be (and have been) successful, provided two key elements are present: an institutional commitment to the program and professional support and recognition of those faculty members who devote their time and energy to the programs.

The problem of students needing remediation in mathematics cannot be wished away and the people who are currently involved in that enterprise deserve both our respect and our gratitude. Teaching remedial mathematics is not easy. It is more enjoyable for most of us to teach abstract algebra to bright students than basic algebra to weak ones. Moreover, it requires fewer teaching skills because the students have more learning skills. But we are failing in our obligations to our students, to mathematics and to society if we do not make substantial efforts to put into place solid, well-conceived, well-supported programs of remediation and to reward our colleagues who design, coordinate and teach in them.

References

1. J. Gardiner, Excellence, Harper, New York, 1961.
2. D. Patterson, Report of Conference on Teaching Remedial Mathematics, April 13–14, 1984, Pomona, CA.

By the amended Theorem above there exists U: $[0,1] \to R$ such that $U(0) = 0$, $U'(x) = 1$ for all rational numbers in $(0,1)$, $U'(x) > 0$ for all x in $[0,1]$ and $U(1) \leqslant \delta$. Thus U is an upper function for f and we have

$$0 \leqslant G\underline{\int_0^1} f(x)\, dx \leqslant G\overline{\int_0^1} f(x)\, dx \leqslant U(1) \leqslant \delta.$$

Since δ was an arbitrary positive number, we have

$$G\underline{\int_0^1} f(x)\, dx = G\overline{\int_0^1} f(x)\, dx = 0,$$

which says that f is G integrable and $G\int_0^1 f(x)\, dx = 0$.

References

1. E. Hewitt and K. Stromberg, Real and Abstract Analysis, Springer, New York, 1965.
2. Y. Katznelson and K. Stromberg, Everywhere differentiable nowhere monotone functions, this MONTHLY, 81 (1974) 349–354.
3. E. J. McShane, Integration, Princeton Univ. Press, NJ, 1944.
4. I. P. Nathanson, Theory of Functions of a Real Variable, Vol. 2, Ungar, New York, Undated.

PROBLEMS AND SOLUTIONS

*For instructions about submitting **proposed** problems for publication in this department see the inside front cover. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the address given on the inside front cover.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

For instructions about submitting solutions of these Elementary Problems, which should be mailed by March 31, 1987, see the inside front cover. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgement).

E 3171. *Proposed by Dean S. Clark, University of Rhode Island.*

In a town where it would require no more than k times as many introductions as there are people to acquaint everyone in the town with everyone else, at least $x\%$ of the population belongs to a certain clique of mutual acquaintances. For a given positive k, find the minimum possible value of $x = x(k)$.

E 3172. *Proposed by Jordi Dou, Barcelona, Spain.*

Let A' (respectively B', C') be the foot of the altitude from vertex A (respectively B, C) in a triangle ABC. Let H be its orthocenter, and M be an arbitrary point of the plane. Prove that the conics $MABA'B'$, $MBCB'C'$, $MCAC'A'$, $MHCA'B'$, $MHAB'C'$, and $MHBC'A'$ have a point other than M in common.

E 3173. *Proposed by H. W. Oliver, Williamstown, MA.*

Let the real-valued function f be differentiable on the open interval (a, b). For $(x, y) \in D = \{(x, y);\ a < x < b,\ a < y < b\}$, define

$$F(x, y) = \frac{f(x) - f(y)}{x - y} \quad \text{for } x \neq y,$$
$$F(x, x) = f'(x).$$

Give a necessary and sufficient condition (in terms of the function f) that

(i) F be differentiable at each point of D;
(ii) F be continuously differentiable on D.

E 3174. *Proposed by Albert Wilansky, Lehigh University.*

Let $a \in cs$ mean that Σa_n converges. Show that

(i) there exists $b \notin cs$ such that $\inf\limits_{a \in cs} \sup\limits_{n} \left|1 - \dfrac{a_n}{b_n}\right| = 0,$

and

(ii) there exists $a \in cs$ such that $\inf\limits_{b \notin cs} \sup\limits_{n} \left|1 - \dfrac{a_n}{b_n}\right| = 0.$

E 3175. *Proposed by C. Olmsted, University of Alaska, Fairbanks.*

Define an integer valued function d_n on the set of permutations π of $\{1, 2, \dots, n\}$ by

$$d_n(\pi) = \sum_{i=1}^{n} |i - \pi(i)|.$$

It is known that the range $(d_n) = \{0, 2, 4, \dots, \lfloor n^2/2 \rfloor\}$ (cf. E 2424, this MONTHLY, 80 (1973), 692 and its solution in 81 (1974), 668–670). Let $E_n(k)$ be the number of elements in $\{\pi,\ d_n(\pi) = 2k,\ \pi$ is even$\}$ and $O_n(k)$ the number of elements in $\{\pi,\ d_n(\pi) = 2k,\ \pi$ is odd$\}$.

Determine the form of the difference $E_n(k) - O_n(k)$.

E 3176. *Proposed by Anatole Beck, The London School of Economics and Political Science, England.*

Let A be a measurable subset of $\mathbb{R}$. Is it true that the interior of $A - A$ (set of differences) either contains 0 or contains nothing?

SOLUTIONS OF ELEMENTARY PROBLEMS

$\dfrac{x^n - y^n}{x - y}$ Integral for All Positive Integral n

E 2998 [1983, 335]. *Proposed by Clark Kimberling, University of Evansville.*

Suppose x and y are complex numbers such that $(x^n - y^n)/(x - y)$ is an integer for some four consecutive positive integers n. Prove it is an integer for all positive integers n.

Solution by A. A. Jagers, Technische Hogeschool Twente, Enschede, Netherlands. Denote

$(x^n - y^n)/(x - y)$ by t_n. Then $(t_n)_{n=0}^{\infty}$ is the unique solution of a difference equation

$$t_{n+2} + bt_{n+1} + ct_n = 0 \quad (n \geqslant 0)$$

with characteristic polynomial $\lambda^2 + b\lambda + c = (\lambda - x)(\lambda - y)$, under the initial conditions $t_0 = 0$, $t_1 = 1$. So it suffices to show that b and c are integers. Assume that $t_n \in \mathbb{Z}$ for $n \in \{m, m+1, m+2, m+3\}$, say. Then

$$c^n = (xy)^n = t_{n+1}^2 - t_{n+2}t_n \in \mathbb{Z} \quad \text{for } n = m \text{ and } n = m+1.$$

Hence $c^{m+1} \in \mathbb{Z}$ and $c \in \mathbb{Q}$, that is, $c \in \mathbb{Z}$. As for b, since $t_{n+2} + bt_{n+1} + ct_n = 0$ for $n = m$ and $n = m+1$,

$$b = (t_m t_{m+3} - t_{m+1}t_{m+2})/c^m$$

by Cramer's rule, and so $b \in \mathbb{Q}$. On the other hand, $t_{m+1} \in \mathbb{Z}$ and, since $c \in \mathbb{Z}$, it easily follows by induction on n, that $t_n = f_n(b)$ for some monic polynomial $f_n(x) \in \mathbb{Z}[x]$ of degree $n - 1$. Hence $b \in \mathbb{Z}$ as well (being a rational algebraic integer).

Also solved by M. Bencze (Romania), D. M. Bloom, R. Breusch, D. M. Broline, P. S. Bruckman, L. A. G. Dresel (United Kingdom), N. J. Fine, I. Gessel, S. Henderson, E. D. Huthnance, L. Kuipers (Switzerland), P. J. Leah and J. B. Wilker (Canada), L. E. Mattics, D. Moews, I. Pranata (student), D. A. Rawsthorne, M. Taboada and C. Zaltzman (Uruguay), A. Tissier (France), E. T. Wong, and the proposer.

Broline, Bruckman and Fine each provided an example showing that the number four in the statement of the problem cannot be replaced by the number three.

Continuous Functions Which Take Some Value an Odd Number of Times

E 3001 [1983, 400]. *Proposed by M. Slater, University of Bristol.*

Let f: $[0, 1] \to \mathbb{R}$ be a continuous function. (a) Show that if f is piecewise strictly monotone, then f takes some value an odd number of times. (b) Show that if f takes every value only finitely often, and $f(0) \neq f(1)$, then f takes some value an odd number of times.

Solution I *by James Propp (student), University of California at Berkeley.* (a) Suppose f is strictly monotone on each of the n intervals $[t_{i-1}, t_i]$, $1 \leqslant i \leqslant n$, with $0 = t_0 < t_1 < \cdots < t_n = 1$. (As is customary, we interpret "piecewise" so that only finitely many pieces are permitted.* See footnote on p. 735.) The set $Y = \{f(t_i)\colon 0 \leqslant i \leqslant n\}$ has at most $n + 1$ distinct elements; label them $y_0 < y_1 < \cdots < y_m$. Set $z_{2i} = y_i$ for $0 \leqslant i \leqslant m$, and pick $z_1, z_3, \ldots, z_{2m-1}$ so that $z_0 < z_1 < z_2 < z_3 < \cdots < z_{2m-1} < z_{2m}$. Let $X_k = \{x\colon f(x) = z_k\}$ and $X = X_0 \cup \cdots \cup X_{2m} = \{x_1, \ldots, x_N\}$, with $0 = x_1 < x_2 < \cdots < x_N = 1$. For $1 \leqslant j \leqslant N$, let k_j be the unique k with $f(x_j) = z_k$. Then $k_1, k_2, \ldots, k_N$ is an integer sequence with k_1 and k_N both even and with $k_j - k_{j+1} = \pm 1$ for $1 \leqslant j < N$. It follows that $N = |X|$ is odd, so that one of the disjoint sets $X_k = f^{-1}(z_k)$ must have an odd number of elements, as claimed.

(b) f is said to have a *proper local maximum* at $x \in [0, 1]$ if and only if there exists a neighborhood V of x such that $f(t) < f(x)$ for all $t \neq x$ in V. It is easy to show (as in [**1**]) that f can have at most countably many such points: if we associate to each such x two rational numbers p, q such that $f(t) < f(x)$ for all $t \neq x$ in $V_x = (p, q)$, then the map $x \mapsto V_x$ is one-to-one with countable range. For the same reason, f has at most countably many proper local minima. We may therefore choose y between $f(0)$ and $f(1)$ so that f has no proper local extrema at any of the elements $x_1 < x_2 < \cdots < x_n$ of $f^{-1}(y)$. Put $x_0 = 0$, $x_{n+1} = 1$, and $I_i = (x_i, x_{i+1})$. Then $f(t) - y$ is nonzero and has constant sign on each I_i, and as a consequence of our choice of y, the sign alternates with respect to i. Since f is negative on I_0 and positive on I_n (or vice versa), n must be odd.

Reference

1. E. E. Posey and J. E. Vaughan, Functions with a proper local maximum in each interval, this MONTHLY, 90 (1983) 281–282.

Editorial Note: In part (a) let C denote the set of all numbers c for which the cardinality of the set $f^{-1}(c)$ is odd (even). It is not too hard to see from this argument that the cardinality of C is either infinite or odd (even).

Solution II *by Mark Bowron* (*student*), *University of Washington*. At a point x in $(0, 1)$, we say f *crosses* the line $y = f(x)$ if in every neighborhood of x, f takes values both less and greater than $f(x)$; if some neighborhood of x lacks this property, we say f *touches* $y = f(x)$ (or has a touch) at x. When all values are taken finitely often, the intermediate-value theorem implies:

(i) If $f(0) \neq f(1)$, then lines parallel to and strictly between the lines $y = f(0)$ and $y = f(1)$ are crossed by f an odd number of times.

(ii) If $f(0) = f(1)$, then lines parallel to but not equal to the line $y = f(0)$ are crossed by f an even number of times.

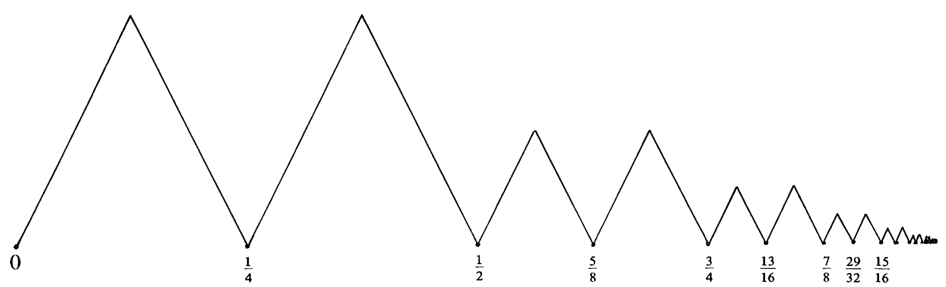
In (a), suppose $f(0) = f(1)$. We can assume $f(0)$ is taken an even number of times, say at $x_1 < \cdots < x_{2n}$. Piecewise strict monotonicity implies that f has an odd number of touches on each subinterval (x_k, x_{k+1}). Since there are an odd number of these subintervals, some line parallel to but not equal to the line $y = f(0)$ is touched by f an odd number of times. Together with (ii), this implies that some value is taken, in total, an odd number of times.

We finish (a) by proving (b). Let E be the set on which f has touches. We claim E is a countable set when all values are taken finitely often.

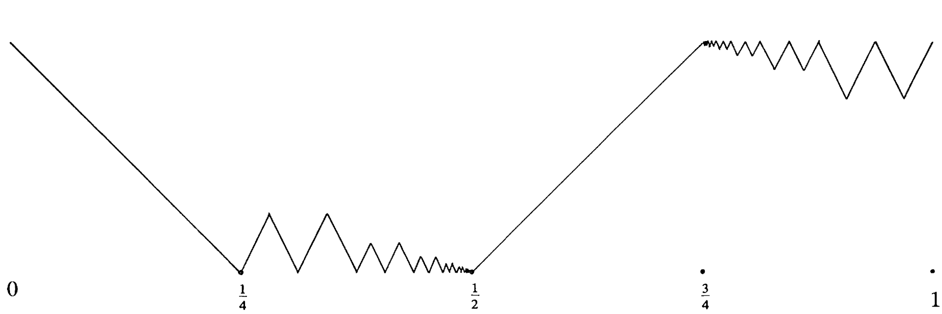
Let $t \in [0, 1]$. Define $g(t, y) = \min\{x: x \geqslant t, f(x) = y\}$ if $y \in f([t, 1])$, $g(t, y) = 1$, otherwise. Note, by the intermediate-value theorem, $g \nearrow$ on $[f(t), \infty)$, and $g \searrow$ on $(-\infty, f(t)]$. Hence, the set of discontinuities of g (as a function of y) is countable. Observe: if $x \in E$, then, for all t on some interval in $[0,1]$, g is discontinuous at $f(x)$. Thus, $E \cap g(t, \mathbb{R})$ is countable for all t. Then

$$E \cap g(T, \mathbb{R}), T = \mathbb{Q} \cap [0,1],$$

*If infinitely many pieces are permitted (and if infinity is considered non-odd), the result is no longer true. A counterexample is



Should it be objected that the use of the degenerate interval $[1, 1]$ as one of the pieces is illegal, one can "remedy" the situation by producing the more complex counterexample



which uses two scaled-down copies of the previous curve (one of them inverted) along with two extra segments.

is also countable. But $E = E \cap g(T,\mathbb{R})$, so E is countable.

RESULT. Uncountably many lines parallel to and strictly between the lines $y = f(0)$ and $y = f(1)$ are never touched by f. Together with (i), this implies that uncountably many values are taken an odd number of times.

Also solved by S. Bäder (West Germany), D. C. Carothers, R. O. Davies (England), R. Goldstein, L. Kuipers (Switzerland),, and the proposer.

Composites of Isotomic / Isogonal Conjugate Transformations

E 3013 [1983, 566]. *Proposed by Stanley Rabinowitz, Digital Equipment Corp.*

Let ABC be a fixed triangle in the plane. Let T be the transformation of the plane that maps a point P into its isotomic conjugate (relative to ABC). Let G be the transformation that maps P into its isogonal conjugate. Prove that the mappings TG and GT are affine collineations (linear transformations).

Solution (adapted from solutions by R. H. Eddy, Memorial University of Newfoundland, and Peter Yff, American University of Beirut). The question is incorrectly phrased (as was noted by H. Guggenheimer and others). Note that the mappings GT and TG are not everywhere well-defined, nor do they always send parallel lines to *parallel* lines; hence they are not affine transformations. However they do send lines to lines where they are defined. This can be seen as follows: If P has trilinear coordinates (x, y, z) relative to triangle ABC, then the isotomic conjugate $T(P)$ has trilinear coordinates proportional to $(b^2c^2x^{-1}, a^2c^2y^{-1}, a^2b^2z^{-1})$, while the isogonal conjugate $G(P)$ has trilinear coordinates proportional to (x^{-1}, y^{-1}, z^{-1}). Thus, if P has trilinear coordinates (x, y, z), then $GT(P)$ has trilinear coordinates proportional to (a^2x, b^2y, c^2z) while $TG(P)$ has trilinear coordinates proportional to $(b^2c^2x, a^2c^2y, a^2b^2z)$. It follows that GT and TG preserve collinearity where they are defined.

Also solved by A. Bondeson (Denmark), R. Cuculière (France), J. Dou (Spain), L. Kuipers (Switzerland), O. P. Lossers (The Netherlands), and the proposer.

Another Sum Identity

E 3022 [1983, 645]. *Proposed by John Sadowsky, Columbia, MD.*

Show that, for any $\alpha > 0$ and any positive integer N,

$$\sum_{k=1}^{N} (-1)^{k-1}\binom{N}{k}\frac{k}{1+(k-1)\alpha} = \prod_{k=1}^{N-1}\left(\frac{k+1}{k+\alpha^{-1}}\right).$$

A number of solvers used the partial fractions technique to confirm the result of this problem. The following solution is representative of that approach.

Solution by O. P. Lossers, Eindhoven University of Technology, The Netherlands. Let $\alpha^{-1} = x$. Then, a partial fractions decomposition of the right-hand side gives

$$\prod_{k=1}^{N-1}\frac{k+1}{k+x} = x\prod_{k=0}^{N-1}\frac{k+1}{k+x} = x\sum_{k=0}^{N-1}\frac{a_k}{k+x},$$

with

$$a_k = N!\prod_{\substack{j=0\\ j\neq k}}^{N-1}\frac{1}{j-k} = \binom{N}{k+1}(k+1)(-1)^k.$$

So

$$\prod_{k=1}^{N-1}\frac{k+1}{k+x}=x\sum_{k=0}^{N-1}(-1)^k\binom{N}{k+1}\frac{k+1}{k+x}=\sum_{k=1}^{N}(-1)^{k-1}\binom{N}{k}\frac{kx}{k-1+x}.$$

Editorial comment. It should be noted that the condition on α can be relaxed to complex $\alpha\neq 0,-n^{-1}$ for $n=1,2,\ldots,(N-1)$.

Also solved by 47 other readers and the proposer.

$\alpha+\beta=1$ and $\alpha^{-1}+\beta^{-1}=1$

E 3029 [1983, 706]. *Proposed by S. W. Golomb, University of Southern California.*

Let α and β be nonzero elements of a field F which satisfy $\alpha+\beta=1$ and $\alpha^{-1}+\beta^{-1}=1$. Prove that $\alpha=\beta^{-1}$ and that $\alpha^6=\beta^6=1$.

Solution I *by Gary F. Birkenmeier, University of Southwestern Louisiana.* The following proposition generalizes problem E 3029 and is an immediate consequence of the lemma. Throughout, R denotes an associative ring with identity.

LEMMA. *Let $\alpha,\beta\in R$ such that $\alpha+\beta=1$.*
(i) *If $\alpha\beta=1$, then $\alpha^3=-1=\beta^3$.*
(ii) *If α and β are invertible, then $\alpha^{-1}+\beta^{-1}=1$ if and only if $\alpha\beta=1$.*

Proof. (i) Since $\beta=1-\alpha$, $\alpha\beta=\alpha-\alpha^2=1$. Consequently,

$$\alpha^3=\alpha^2\alpha=(\alpha-1)\alpha=\alpha^2-\alpha=-1.$$

Similarly, $\beta^3=-1$.

(ii) If $\alpha^{-1}+\beta^{-1}=1$, then $\alpha\beta=\alpha(\alpha^{-1}+\beta^{-1})\beta=\beta+\alpha=1$. If $\alpha\beta=1$, then $\alpha=\beta^{-1}$ and $\beta=\alpha^{-1}$. Thus,

$$1=\alpha+\beta=\beta^{-1}+\alpha^{-1}=\alpha^{-1}+\beta^{-1}.$$

PROPOSITION. *Let $\alpha,\beta\in R$ such that α and β are invertible and $\alpha+\beta=1=\alpha^{-1}+\beta^{-1}$. Then $\alpha=\beta^{-1}$ and $\alpha^6=1=\beta^6$.*

Solution II *by Fergus J. Gaines, University College Dublin, Ireland.* Let A be a 2×2 matrix with eigenvalues α and β. Then A^{-1} has eigenvalues α^{-1} and β^{-1} and trace $A=1=$ trace A^{-1}. The characteristic polynomial of A is

$$\lambda^2-(\operatorname{trace}A)\lambda+(\det A)=\lambda^2-\lambda+d$$

where $d=\det \mathrm{A}=\alpha\beta$. The Hamilton-Cayley theorem implies that $A^2-A+dI=0$. Multiplying by $d^{-1}A^{-2}$ gives

$$A^{-2}-d^{-1}A^{-1}+d^{-1}I=0.$$

Since A^{-1} has distinct eigenvalues, its characteristic polynomial is $\lambda^2-d^{-1}\lambda+d^{-1}$. But trace $A^{-1}=1$. So $d=1$. Thus, $\alpha\beta=1$, and $\beta=\alpha^{-1}$. Since α is an eigenvalue of A, $\alpha^2-\alpha+1=0$. So

$$\alpha^3+1=(\alpha+1)(\alpha^2-\alpha+1)=0,$$

and $\alpha^3=-1$. Thus, $\alpha^6=1$ and also $\beta^6=1$.

Comment by Fred Safier, City College of San Francisco. We may note that the system $\alpha+\beta=1$, $\alpha^{-1}+\beta^{-1}=1$ has no solutions in Q or in R. In C, of the possible sixth roots of unity, α and β can be chosen only from

$$\frac{1+i\sqrt{3}}{2},\quad \frac{1-i\sqrt{3}}{2}.$$

In fact, in any field the conditions $z^6 = 1, z \neq 1$ suggest we look for solutions satisfying

$$z + 1 = 0, z^2 - z + 1 = 0, \quad \text{or} \quad z^2 + z + 1 = 0.$$

In Z_p we find, as in C, that all solutions satisfy $z^2 - z + 1 = 0$, i.e., $z = (1 \pm \sqrt{-3})/2$. Therefore, there are solutions in any field Z_p in which $(1 \pm \sqrt{-3})/2$ is defined. This includes, but is not limited to, all fields Z_p where p is of the form $n^2 + 3$. For example, in Z_3 we have solutions $\alpha = \beta = 2$, in Z_7 we have solutions $\alpha, \beta = 3$ or 5, and in Z_{67} we have solutions $\alpha, \beta = 38$ or 30. $(1 \pm \sqrt{-3})/2$ is also defined, however, for example, in such other fields as Z_{13}, where $\alpha, \beta = 4$ or 10, and Z_{31} where $\alpha, \beta = 26$ or 6. $(1 \pm \sqrt{-3})/2$ is not defined in Z_2, Z_5, Z_{11}, Z_{17}, Z_{23}, or Z_{29}, and there are no solutions in these fields.

There were 165 correct solutions and 6 incorrect solutions received. Many solvers provided generalizations other than those given in the printed solutions above.

A Twice Differentiable Function f with f, f', f'' Increasing

E 3033 [1984, 57]. *Proposed by M. McAsey, Bradley University, and L. A. Rubel, University of Illinois.*

Let f be a twice differentiable function on $(-\infty, \infty)$ with f, f', f'' increasing. Fix numbers a and b with $-\infty < a \leqslant b < \infty$. For each $x > 0$, define $\xi = \xi(x)$ so that

$$\frac{f(b+x) - f(a-x)}{b - a + 2x} = f'(\xi)$$

by the Mean Value Theorem. Prove that $\xi(x)$ is an increasing function of x. Can the hypothesis f'' increasing be replaced with f'' positive?

Composite solution based on solutions by Vania D. Mascioni (student), ETH, Zurich, and Renato Spigler, Courant Institute, New York. The equation

$$\xi(x) = (f')^{-1}\left(\frac{f(b+x) - f(a-x)}{b - a + 2x}\right)$$

makes it clear that ξ is differentiable. Differentiating the equation defining $f'(\xi)$ yields

$$\text{(1)} \qquad \frac{f'(b+x) + f'(a-x) - 2f'(\xi)}{(b - a + 2x)} = f''(\xi) \cdot \xi'(x).$$

Since f' is increasing, $f'' \geqslant 0$. Thus $\xi'(x) > 0$ would follow from (1) provided that

$$\text{(2)} \qquad \frac{f'(b+x) + f'(a-x)}{2} > f'(\xi).$$

But f'' increasing implies f' convex, hence (2) holds.

The condition that f'' increases can not be replaced with $f''(x) > 0$. Consider

$$f(x) = \frac{\pi}{2}x + \int_0^x \arctan(t)\, dt.$$

Then

$$f'(x) = \frac{\pi}{2} + \arctan(x), f''(x) = 1/(1 + x^2),$$

so that f and f' are increasing, $f''(x) > 0$, but f'' decreases on $\mathbb{R}^+$. Setting $a = b = 0$, we find that the equation defining $\xi(x)$ becomes

$$\frac{\pi}{2} = \frac{\pi}{2} + \arctan \xi(x),$$

hence $\xi(x) \equiv 0$.

Also solved by K. F. Anderson (Canada), D. M. Bloom, D. Dugosija (Yugoslavia), T. Jager, W. Janous (Austria), R. T. Koether and W. H. Meyers, D. Landsay, O. P. Lossers (The Netherlands), J. -M. Monier (France), W. A. Newcomb, S. Noltie, P. J. Paul (Spain), S. Phillip, R P. Savage, Jr., A. Villani (Italy), and the proposers.

Bloom, Jager, Noltie, Phillip, and Villani point out that the hypothesis f increasing is not needed. Villani proves the following more general result: Let $f\colon \mathbb{R} \to \mathbb{R}$ be continuously differentiable. For $a, b \in \mathbb{R}$, $a \leqslant b$, let $F_{a,b}\colon \mathbb{R}^+ \to \mathbb{R}$ be defined by

$$F_{a,b}(x) = \frac{f(b+x) - f(a-x)}{b - a + 2x}, \quad x \in \mathbb{R}^+;$$

then f' is (strictly) convex if and only if each $F_{a,b}$ is (strictly) increasing.

Limit of a Modified Sequence

E 3034 [1984, 58]. *Proposed by David Cox, Battelle Memorial Institute, Columbus, OH.*

Let $0 < x_0 < 1$ and $x_{n+1} = x_n - x_n^2$, for $n = 0, 1, 2, \ldots$. Prove that

$$\lim_{n\to\infty} \frac{n(1 - nx_n)}{\log n} = 1.$$

Solution by O. P. Lossers, Eindhoven University of Technology, The Netherlands. Substitution of $y_n = x_n^{-1}$ in $x_{n+1} = x_n - x_n^2$ yields the equivalent recurrence relation

$$(1) \qquad y_{n+1} = y_n^2/(y_n - 1) = y_n + 1 + \frac{1}{y_n - 1},$$

where

$$(2) \qquad y_n > 1.$$

It follows from (1) and (2) and a trivial induction argument that for all n

$$(3) \qquad y_n \geqslant n.$$

From (3) and (1) one finds the existence of a constant C such that for all n

$$(4) \qquad \begin{aligned} y_n &\leqslant y_{n-1} + 1 + \frac{1}{n-1} \leqslant y_{n-2} + 2 + \frac{1}{n-1} + \frac{1}{n-2} \leqslant \cdots \\ &\leqslant y_1 + n + \sum_{i=1}^{n-1} \frac{1}{i} \leqslant n + \log n + C. \end{aligned}$$

This in turn proves the existence of a constant C' such that for all n

$$(5) \qquad \begin{aligned} y_n &\geqslant y_{n-1} + 1 + \frac{1}{n - 1 + \log n + C} \geqslant \cdots \\ &\geqslant y_1 + n + \sum_{i=1}^{n-1} \frac{1}{i + \log(i+1) + C} \geqslant n + \log n + C'. \end{aligned}$$

It follows from (4) and (5) that on the one hand

$$\frac{n(1 - nx_n)}{\log n} = \frac{n(y_n - n)}{y_n \log n} \leqslant \frac{n(\log n + C)}{(n + \log n + C')\log n} = 1 + o(1) \quad (n \to \infty),$$

while on the other hand

$$\frac{n(1 - nx_n)}{\log n} = \frac{n(y_n - n)}{y_n \log n} \geqslant \frac{n(\log n + C')}{(n + \log n + C)\log n} = 1 + o(1) \quad (n \to \infty).$$

We conclude that

$$\lim_{n\to\infty} \frac{n(1-nx_n)}{\log n} = 1.$$

Also solved by M. Bencze (Romania), L. M. Christophe, Jr., R. Cuculière (France), J. Deutsch, D. Dugošija (Yugoslavia), S. Friedlander, L. Gerber, S. Heller, W. Janous (Austria), J. -M. Monier (France), W. A. Newcomb, M. S. Perkins, S. Philipp, A. Smuckler (Israel), A. Stenger, R. Szwarc (Poland), M. Vowe (Switzerland), P. Y. Wu (Republic of China), and the proposer.

Congruence mod p^k of Sums of Binomial Coefficients

E 3035 [1984, 140]. *Proposed by L. E. Mattics, University of South Alabama.*

Let p be an odd prime. For $1 \leqslant m \leqslant p-1$ and $0 \leqslant i \leqslant p-1$ set

$$s(m,i) = \binom{mp}{i} - \binom{mp}{i+p} + \binom{mp}{i+2p} - \cdots .$$

Prove

$$s(m,0) \equiv 0 \pmod{p^{m+1}} \quad \text{for } 1 \leqslant m \leqslant p-2,$$

and

$$s(m,i) \equiv p^m(-1)^{i+1}m!i^{-m} \pmod{p^{m+1}} \quad \text{for } i > 0.$$

($s(p-1,0) \equiv p^{p-1} (\operatorname{mod} p^p)$ is established in E 2685 [1977, 820; 1979, 131–132].)

Solution by the proposer. $s(1,0)=0$. Using Wilson's Theorem, we see that, for $1 \leqslant i \leqslant p-1$,

$$s(1,i) \equiv \binom{p}{i} \equiv p(-1)^{i+1}i^{-1} \pmod{p^2}.$$

Suppose we have proved the statement for all m with $1 \leqslant m \leqslant n < p$. Then because $\binom{a}{b} = \sum_{i=0}^{k}\binom{a-k}{b-i}\binom{k}{i}$,

$$\begin{aligned} s(n+1,0) &= \sum_{r=0}^{n+1}(-1)^r \sum_{j=0}^{p}\binom{np}{(p-j)+(r-1)p}\binom{p}{j} \\ &= \sum_{r=0}^{n+1}(-1)^r\left(\binom{np}{rp}\binom{p}{0} + \binom{np}{(r-1)p}\binom{p}{p}\right) \\ &\quad + \sum_{j=1}^{p-1}\binom{p}{j}\sum_{r=1}^{n}(-1)^r\binom{np}{p-j+(r-1)p} \\ &= 0 - \sum_{j=1}^{p-1} s(1,j)s(n,p-j). \end{aligned}$$

So, by the finite induction hypothesis,

$$\begin{aligned} s(n+1,0) &\equiv -\sum_{j=1}^{p-1} p(-1)^{j+1}j^{-1}p^n(-1)^{p-j+1}n!(p-j)^{-n} \\ &\equiv -p^{n+1}n!\sum_{j=1}^{p-1}(p-j)^{-(n+1)} \pmod{p^{n+2}}. \end{aligned}$$

Thus, if $n+1 = p-1$, we have

$$s(p-1,0) \equiv p^{p-1} \pmod{p^p};$$

and if $n + 1 < p - 1$, then

$$s(n+1,0) \equiv 0 \pmod{p^{n+2}}.$$

Now suppose $1 \leqslant i \leqslant p - 1$. Then

$$\begin{aligned} s(n+1,i) &= \sum_{r=0}^{n}(-1)^r \sum_{j=0}^{p}\binom{np}{i-j+rp}\binom{p}{j} \\ &= \sum_{r=0}^{n}(-1)^r \sum_{j=0}^{i}\binom{np}{i-j+rp}\binom{p}{j} \\ &\quad + \sum_{r=1}^{n}(-1)^r \sum_{j=i+1}^{p}\binom{np}{i-j+rp}\binom{p}{j} \\ &= s(1,i)s(n,0) + \sum_{j=1}^{i-1} s(1,j)s(n,i-j) - \sum_{j=i+1}^{p-1} s(1,j)s(n,p+i-j). \end{aligned}$$

Now by induction,

$$\begin{aligned} s(n+1,i) &\equiv 0 + \sum_{j=1}^{i-1} p^{n+1}(-1)^{j+1}(j^{-1})(-1)^{i-j+1}n!(i-j)^{-n} \\ &\quad - \sum_{j=i+1}^{p-1} p^{n+1}(-1)^{j+1}(j^{-1})(-1)^{i-j+p+1}n!(i-j)^{-n} \\ &\equiv (-1)^i p^{n+1} n! \sum_{\substack{j=1 \\ j\neq i}}^{p-1} \left(j(i-j)^n\right)^{-1} \pmod{p^{n+2}}. \end{aligned}$$

Since

$$\left(1-(1-(j/i))^n\right)/(j/i) = (1-(j/i))^{n-1} + (1-(j/i))^{n-2} + \cdots,$$

$$j^{-1}(i-j)^{-n} = i^{-n}j^{-1} + i^{-n}(i-j)^{-1} + i^{-(n-1)}(i-j)^{-2} + \cdots + i^{-1}(i-j)^{-n},$$

so

$$\begin{aligned} s(n+1,i) &\equiv (-1)^i p^{n+1} n!\left(i^{-n}\left(i^{-1} + \sum_{\substack{j=1 \\ i\neq j}}^{p-1} j^{-1}\right) + \sum_{r=1}^{n} i^{-(n+1-r)}\left(i^{-r} + \sum_{j=1}^{p-1}(i-j)^{-r}\right)\right. \\ &\qquad \left. -(n+1)i^{-(n+1)}\right) \equiv (-1)^{i+1}(n+1)!\,i^{-(n+1)}p^{n+1} \pmod{p^{n+2}} \end{aligned}$$

because $\sum_{j=1}^{p-1} j^{-r} \equiv 0 \pmod{p}$ for $1 \leqslant r \leqslant p - 2$.

Now Even Better Known

E 3038 [1984, 140]. *Proposed by Tetsundo Sekiguchi, University of Arkansas.*

Let ABC be a triangle in the plane. Show that

$$\sin A + \sin B + \sin C \leqslant \frac{3\sqrt{3}}{2} \text{ and } \sin A \sin B \sin C \leqslant \frac{3\sqrt{3}}{8}.$$

Solution I *by Stan Philipp, Santa Maria, California.* Because sine is concave on $[0, \pi]$, we have

$$\frac{\sin A + \sin B + \sin C}{3} \leqslant \sin\left(\frac{A + B + C}{3}\right) = \sin\left(\frac{\pi}{3}\right) = \frac{\sqrt{3}}{2}.$$

By the arithmetic mean-geometric mean inequality and the previous statement, we have

$$\sin A \sin B \sin C \leqslant \left(\frac{\sin A + \sin B + \sin C}{3}\right)^3 \leqslant \left(\frac{\sqrt{3}}{2}\right)^3.$$

Solution II *by Graham Lord, Princeton, New Jersey.* With no loss of generality (because of similarity), consider only those triangles whose circumcircle has diameter 1. Then by the Sine Rule:

$$\sin A + \sin B + \sin C = a + b + c.$$

But the perimeter of a circumscribed polygon is a maximum when a polygon is regular, and is equal to $3\sqrt{3}/2$, in the case of a triangle. In the same family of triangles the product of sines equals abc which by the arithmetic mean-geometric mean inequality is less than or equal to $((a + b + c)/2)^{1/3} = 3\sqrt{3}/8$.

Comment by H. Guggenheimer. The two inequalities are very old and have been rediscovered many times; cf. K. P. Williams, this MONTHLY, 44 (1937) 579–583; E 1644, this MONTHLY, 71 (1964) 915–916; my *Plane Geometry and Its Groups*, problems 7–3 #9, 11, and many more.

As far as I can see, the inequalities are due to G. Berkhan, *Zur projektivischen Behandlung der Dreiecksgeometrie*, Arch. Math. Phys., (3)11(1907) 1–31. The beautiful proof is reproduced in: G. Berkhan-W. Fr. Meyer, *Neuere Dreiecksgeometrie, Encyclopaedie der mathematischen Wissenschaften*, III/1, p. 1194–6, note 78a.

Also solved by 93 other readers and the proposer.

An Inequality

E 3040 [1984, 203]. *Proposed by J. L. Brenner, Palo Alto, CA, and W. A. Newcomb, Lawrence Livermore National Laboratory.*

Let

$$M_1 = \left[\frac{1}{2}\left(a_1 a_2^{\varepsilon} + a_1^{\varepsilon} a_2\right)\right]^{1/(1+\varepsilon)}, \qquad M_2 = \left[\frac{1}{2}\left(a_1^{\varepsilon} + a_2^{\varepsilon}\right)\right]^{1/\varepsilon},$$

where $\varepsilon \neq 0, -1$. For what values of ε is it true for all $a_1 > a_2 > 0$ that $M_1 > M_2$, and for what values is the opposite true?

Solution by Stan Philipp, Pennsylvania State University at Mont Alto. The answer is that

(a) if $\varepsilon < -1$, then $M_1 < M_2$;
(b) if $-1 < \varepsilon < 0$, then $M_1 > M_2$;
(c) if $\varepsilon \geqslant \frac{1}{3}$, then $M_1 < M_2$;
(d) if $0 < \varepsilon < \frac{1}{3}$, then neither $M_1 > M_2$ nor $M_1 < M_2$ holds for all $a_1 > a_2 > 0$.

Proof. Define, for $0 < x \leqslant 1$,

$$\phi(x) = \frac{1}{1+\varepsilon}\log\frac{1}{2}(x + x^{\varepsilon}) - \frac{1}{\varepsilon}\log\frac{1}{2}(1 + x^{\varepsilon}). \tag{1}$$

The relevance of ϕ to our problem is that if $x = a_2/a_1$, then

$$\phi(x) = \log M_1 - \log M_2.$$

We observe that

$$\phi(1) = 0 \tag{2}$$

and

$$\phi'(x) = \frac{x^{1-\varepsilon} - x^{\varepsilon} + \varepsilon(1-x)}{(1+\varepsilon)(x+x^{\varepsilon})(x+x^{1-\varepsilon})}. \tag{3}$$

Argument for (a) and (b): Let $\varepsilon < 0, 0 < x < 1$. Then the numerator in (3) is negative and the denominator has the same sign as $1+\varepsilon$. It follows that $\phi'(x) > 0$ if $\varepsilon < -1$ and $\phi'(x) < 0$ if $-1 < \varepsilon < 0$; now (a) and (b) are immediate by (2).

Argument for (c): Let $\varepsilon \geqslant \frac{1}{3}, 0 < x < 1$. Then the numerator in (3) is (termwise) $\geqslant x^{2/3} - x^{1/3} + \frac{1}{3}(1-x) = \frac{1}{3}(1-x^{1/3})^3 > 0$, so $\phi'(x) > 0$, and (c) now follows from (2).

Argument for (d): Suppose $0 < \varepsilon < \frac{1}{3}, 0 < \mathrm{x} < 1$. By (1), we have

$$\phi(x) < 0 \text{ for sufficiently small } x. \tag{4}$$

Calculations show that $\phi(1) = \phi'(1) = 0$ and $\phi''(1) = (1-3\varepsilon)/4(1+\varepsilon) > 0$; it follows that

$$\phi(x) > 0 \text{ for } x \text{ sufficiently close to } 1. \tag{5}$$

By (4) and (5), the argument for (d) is complete.

Also solved by Aage Bondesen (Denmark), T. Jager, Walther Janous (Austria), Robert E. Shafer, and the proposers.

ADVANCED PROBLEMS

For instructions about submitting solutions of these Advanced Problems, which should be mailed by March 31, 1987, see the inside front cover. The solver's full post-office address should be on each sheet.

6529. *Proposed by O. Hájek, Case Western Reserve University.*

Let $f_1, f_2, \ldots$ be a sequence of functions from R^1 to R^1 with common Lipschitz constant λ. Show that each continuous function f with

$$(\text{graph } f) \subseteq \bigcup_{k=1}^{\infty} (\text{graph } f_k)$$

also has Lipschitz constant λ. (The condition on the graph of f may be restated as follows: for each real x there is a k with $f(x) = f_k(x)$.) Is this true if the Lipschitz condition is replaced by a Hölder condition of order α for some fixed α with $0 < \alpha < 1$?

6530. *Proposed by Wolfgang Walter, Universität Karlsruhe, Federal Republic of Germany.*

Let f be a real-valued continuous function on $(0,1)$, and $\{a_n\}, \{b_n\}$ two sequences of real numbers such that $a_n \neq b_n$, $n = 1, 2, 3, \ldots$, and

$$\lim a_n = \lim b_n = 0.$$

Assume that for each x in $(0,1)$ there exists an $N = N(x)$ for which

$$f(x+a_n) = f(x+b_n), \quad n \geqslant N(x).$$

Must f be a constant function if $0 < b_n < a_n$?

*What if $b_n = -a_n$?

SOLUTIONS OF ADVANCED PROBLEMS

Similar Yet Different

6489 [1985, 148]. *Proposed by Eliot Jacobson, Ohio University.*

Suppose that E and F are fields such that the additive groups $(E, +), (F, +)$ are isomorphic, and the multiplicative groups $(E^*, \cdot), (F^*, \cdot)$ are isomorphic. Are E and F isomorphic as fields?

Solution by J. Bajger, Z. L. Dou, A. Hicks, and R. Thomas (Queens College Honors Algebra Students), Flushing, New York. If E and F are finite extension fields of the rationals Q, then their additive groups are isomorphic if and only if E and F are isomorphic as Q-vector spaces; that is, if and only if E and F have the same degree over Q.

Suppose further that E is the quotient field of a unique factorization domain R which has denumerably many (non-associate) primes $\rho_1, \rho_2, \dots$ and unit group U. Then every nonzero element x of E has a unique representation in the form

$$x = u\prod \rho_i^{e_i},$$

where u is in U and the e_i are integers, all but finitely many of which are zero. Clearly the map

$$x \to (u, e_1, e_2, \dots)$$

provides an isomorphism of E^* with the direct sum of U and denumerably many copies of the integers Z.

If we take for example the fields $E = Q(\sqrt{-2})$ and $F = Q(\sqrt{-7})$, their respective unit groups are $\{\pm 1\}$ while their respective rings of integers are Euclidean and thus have the unique factorization property. Hence E^* and F^* are isomorphic, as are $(E, +)$ and $(F, +)$. However, E and F cannot be isomorphic as fields since $x^2 + 2$ has two roots in E but none in F.

Some readers pointed out that the problem is essentially solved in L. Fuchs, *Infinite Abelian Groups*, vol 2, p. 314, via a theorem of T. Skolem, Norske Vid. Selsk. Forh. 2 (1947), 4–7. Others parlayed the impossibility of solving all fifth degree equations by radicals into more exotic counterexamples. An approach rather different from the above was to construct pairs of fields satisfying both isomorphisms, but with different transcendence degrees. One such pair is $Z_2(x)$ and $Z_2(x, y)$, both extensions of Z_2, the field of two elements. Another is provided by the algebraic closures of $Q(x)$ and $Q(x, y)$ where Q is the field of rationals.

The proposer remarks that the number field examples are of interest in connection with Dedekind zeta functions (see Robert Perlis, *On the equation* $\zeta_K(s) = \zeta_{K'}(s)$, J. Number Theory, 9 (1977), 342–360).

Also solved by 21 others and the proposer.

More like $\zeta(2[n/2])$ than $\zeta(2n)$

6490 [1985, 148]. *Proposed by J. Sutherland Frame, Michigan State University.*
Show that

$$\sum_{k=1}^{\infty} (k^2 + k)^{-3} = 10 - \pi^2,$$

and more generally that

$$(-1)^{n-1} \sum_{k=1}^{\infty} (k^2 + k)^{-n} = \binom{2n-1}{n-1} + \sum_{j=1}^{[n/2]} \binom{2n-1-2j}{n-1} B_{2j}(-4\pi^2)^j/(2j)!,$$

where $B_2, B_4, B_6, \dots$ are the Bernoulli numbers $1/6, -1/30, 1/42, -1/30, 5/66, \dots$.

Solution by William A. Newcomb, Lawrence Livermore National Laboratory. Define

$$I_{mn} = \sum_{k=1}^{\infty} \frac{1}{k^m(k+1)^n}, \quad m, n \geqslant 0, m+n \geqslant 2.$$

Then we have the recurrence relation

$$\begin{aligned} I_{m+1,n+1} &= \sum_{k=1}^{\infty} \frac{1}{k^m(k+1)^n}\left(\frac{1}{k} - \frac{1}{k+1}\right) \\ &= I_{m+1,n} - I_{m,n+1}, \quad m, n \geqslant 0, m+n \geqslant 1. \end{aligned}$$

The initial values are

$$I_{11} = 1, \quad I_{n0} = \zeta(n), \quad I_{0n} = \zeta(n) - 1, \quad n \geqslant 2.$$

By induction on m and n, it is easily shown that

$$I_{n1} = (-1)^n\left[-1 + \sum_{k=2}^{n} (-1)^k \zeta(k)\right], \quad n \geqslant 2,$$

$$I_{1n} = n - \sum_{k=2}^{n} \zeta(k), \quad n \geqslant 2,$$

and

$$\begin{aligned} I_{mn} = (-1)^m\Bigg[-\binom{m+n-1}{n-1} &+ \sum_{k=2}^{m} (-1)^k \binom{m+n-k-1}{n-1}\zeta(k) \\ &+ \sum_{k=2}^{n} \binom{m+n-k-1}{m-1}\zeta(k)\Bigg], \quad m, n \geqslant 2. \end{aligned}$$

(For this, we use

$$\binom{p}{q} = \binom{p-1}{q} + \binom{p-1}{q-1}$$

and

$$\binom{p}{p} + \binom{p+1}{p} + \cdots + \binom{p+q}{p} = \binom{p+q+1}{p+1}.)$$

For $m = n$ the terms with odd k cancel, so

$$I_{nn} = (-1)^n\left[-\binom{2n-1}{n-1} + 2\sum_{j=1}^{[n/2]} \binom{2n-2j-1}{n-1}\zeta(2j)\right], \quad n \geqslant 2.$$

In particular,

$$I_{22} = -3 + 2\zeta(2), \quad I_{33} = 10 - 6\zeta(2) = 10 - \pi^2.$$

The general result follows from the well-known formula

$$\zeta(2j) = -\frac{(-4\pi^2)^j}{2(2j)!}B_{2j}.$$

H. Roelants (Belgium) writes that the result exists "in disguised form" in J. W. L. Glaisher, *Summation of certain numerical series*, Messenger Math., 1913 (42), 19–34. He adds that related material and further references can be found in Bruce Berndt, *Ramanujan's Notebooks Part I*, Springer-Verlag, 1985 (in particular, see Chap. 7, Entry 22 and Chap. 9, Entry 35).

Also solved by twenty-nine other readers and the proposer; a partial solution was also received.

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER

Linear Algebra. By P. G. Kumpel and J. A. Thorpe. Saunders College Publishing, 1983, ix + 353.
Linear Algebra: *An Introductory Approach*. By Charles W. Curtis. Springer-Verlag, 1984, x + 337.
Linear Algebra with Applications (Second Edition). By Jeanne Agnew and Robert C. Knapp, Brooks/Cole Publishing Company, 1983, xii + 361.
Linear Algebra, second edition. By Larry Smith. Springer-Verlag, 1984, vii + 356.
Applications of Linear Algebra, third edition. By Chris Rorres and Howard Anton. John Wiley & Sons, Inc., 1984, ix + 341.

PATRICIA C. KENSCHAFT
Department of Mathematics, Montclair State College, Upper Montclair, NJ 07043

How many different courses march in the parade called "Linear Algebra"? Do we already offer undergraduates "Soft Linear Algebra," "Compromise Linear Algebra," and "Highbrow Linear Algebra"? Should we? What is an "application"?

While the extraordinary diversity of American education may be a source of national strength, it can generate disaster for individual Americans. If a student takes one course called "Linear Algebra" as an undergraduate and a graduate department assumes more advanced material has been learned, the student can be utterly lost. This would seem to be occasionally inevitable in a nation that offers at least three distinct college courses called "Linear Algebra." It is time that American mathematicians be more explicit with ourselves and our fellow citizens.

Complicating the problem is the ambiguity of the omnipresent word "applications," since it apparently means nothing more than a topic outside the mainstream of the book in which it appears. Unlike the other four books in this set, Rorres and Anton is intended as a companion book for another text that provides enriching "applications." It does not attempt to present a cohesive course. Its "applications" include economic models, forest management and harvesting of animal populations—but also plane geometry and linear programming! Agnew and Knapp (henceforth A & K) places its applications, as has been the fashion in recent years, in sections at the end of each chapter. These applications include an explanation of BASIC programs for Gauss-Jordan row reduction and Gram-Schmidt orthogonalization. The book presents many interesting non-mathematical applications as well, but these generally are not so labeled in the table of contents, but are called merely "examples." Curtis's introduction manifests yet another attitude, touting as "applications" its chapters on finite symmetry groups, the exponential of a matrix, and the composition of quadratic forms. This inconsistent use of a common word might be considered a diabolical flaw in a discipline that prides itself on careful definitions. It may, however, indicate a laudable, if ironic, flexibility among mathematicians, resulting not only from pressures of the market-place but also our own desire to be widely useful.

Although it matters little whether most topics are called "applications," such a designation of "Systems of Linear Equations" by Curtis demonstrates that the book is far more advanced than A & K or Kumpel and Thorpe (henceforth K & T). Both of these, following the historical development of linear algebra, use Systems of Linear Equations as a fundamental introductory topic. The other crucial indicator of the level of the course is linear independence. Both Curtis and Smith introduce this concept about a tenth of the way through their pages. In contrast, A & K postpones the topic of linear independence to the middle of the book, and K & T introduces it only somewhat sooner.

The four books other than Rorres and Anton can be linearly ordered in the zig-zag sense that human creations can be ordered. A & K represents a large class of texts that, with some modification, could be used just as easily before a student studies calculus, perhaps long before.

Although this particular book concludes with solving systems of differential equations by diagonalization of matrices, its first half is not much more difficult than traditional high school algebra. This is accomplished by avoiding linear independence while presenting diagonalization of matrices. In doing so it covers material crucial not just to mathematics but also to computer science, economics, and physics, without undue challenge to the brain.

Apparently, something similar to the one-semester non-calculus course frequently offered to administrative science majors is appropriate to a far larger audience, perhaps including beginning mathematics majors. Motivating matrix diagonalization by such applications as Markov chains and animal population stabilization (instead of differential equations), adding related topics such as linear programming, and requiring some elementary computer programming could result in an intellectually respectable package that does not require total digestion of linear independence for its consumption. Early in my career I was shocked to receive a memo from a since-retired colleague saying that his linear algebra students had become lost in a "sea of vector spaces" and proposing that the entire sea be circumvented. The suggestion ceased to seem heretical only after I first tried to steer my own beginning linear algebra students into abstract vector spaces. I now believe that given the preparation of many current high school graduates, a first course in linear algebra might concentrate on matrix manipulations and applications, motivating an understanding of linear independence whenever possible, but not requiring it for completion of the course. It might be called "Linear Algebra: Matrix Theory."

Although such a course, I suspect, would receive college credit for a few more decades, the question remains open as to how hard we should try to push it to the pre-college level. Now may be the time to press for more substance in the mathematics requirement for elementary school teachers. Instead of a vague review, why shouldn't they broaden their mathematical horizons by taking Matrix Theory? Such a course would review prior topics while demonstrating the utility of mathematics in a variety of non-mathematical fields. Our experience with administrative science majors indicates that this course and one in basic probability and statistics are *not* too difficult to require from non-technical majors. In the next decade many new elementary teachers must be prepared to teach the (now preschool) "baby boomlet." If we set precedents soon, while the supply of aspiring elementary teachers exceeds the demand, our rewards in the mathematical literacy of our countryfolk three decades from now might be substantial.

To assert that a course is worthy of college credit is not to aver that it shouldn't be offered sooner. Students who actually learn Algebra 1 as scheduled might undertake Matrix Theory under the already expansive umbrella called Algebra 2. Even better, Matrix Theory might follow an early Algebra 1 for gifted middle school youngsters, especially if the dying SMSG (now in the form of Addison-Wesley's Unified Mathematics) finally goes out of print. In most nations students learn our Algebra 1 around age eleven or twelve. Surely there are many American youngsters who want and need intellectual fare more nearly comparable to that offered to their peers in other countries. Preparing their teachers to share mathematical concepts before their intellectual receptivity has dissipated would help prevent our own progeny from becoming bored with school before they enter our collegiate halls.

Mathematics majors who took Matrix Theory before or concurrently with calculus would be prepared for a course along the lines of Smith or Curtis before college graduation. It might be called "Linear Algebra: Vector Space Theory," and it would assume a certain mathematical devotion from its students with the accompanying willingness to apply themselves to rigorous abstractions. Neither of these books includes applications to non-mathematical topics, nor should they. One can conceive of a world in which the charm and pervasiveness of mathematics was so overwhelming that academia had such fragments as the Department of Integers (listing chemical elements, birds, and African tribes), the Department of Matrices (displaying interrelationships of ecological chains, families, and crystals), the Department of Random Processes (exploring Heisenberg's principle, human learning, and the results of advertising), and the Department of the Real Numbers (considering moving objects, changing temperatures, and growth of trees). In such a world mathematics itself would be interdisciplinary, but that would not necessarily make it more

appealing. Nor, in our world (where many college students come from homes in which Leontief matrices are as arcane as the Spectral Theorem) do non-mathematical applications necessarily render mathematics more fascinating to its students.

Vector Theory itself provides sufficient challenge and intrinsic reward if it reaches the Spectral Theorem and the Principal Axis Theorem, both crucial for graduate work in many mathematical fields. Smith proves the Spectral Theorem in the form that any self-adjoint linear transformation over a finite dimensional vector space can be represented by a diagonal matrix using a suitably chosen orthonormal basis, where a self-adjoint linear transformation T is defined to be one for which $(Tx, y) = (x, Ty)$ for all vectors x and y in the vector space. The Principal Axis Theorem follows quickly in the form that asserts that every quadratic function q from the vector space to the real numbers (that is, a function of the form $q(x) = (x, Tx)$ for some self-adjoint linear transformation T) can be expressed as

$$q(x) = \lambda_1 c^2 + \lambda_2 c^2 + \cdots + \lambda_n c^2,$$

where $x = c_1x_1 + c_2x_2 + \cdots + c_nx_n$ and $\{x_i\}$ is a judiciously chosen orthonormal basis for the original vector space consisting of eigenvectors of T, with corresponding eigenvalues λ_i.

Smith may be read as a good novel whose plot is known but whose beauty may be savored as it unfolds. The author gives us signposts to mark our way. For example, after proving that the only eigenvalue of a nilpotent operator (one for which some power is the zero operator) is zero, he adds, "In a very strong sense this is the reason why not all transformations can be diagonalized." An idealistic reader might wish for the polish of a fastidious American editor who had removed occasional surface bumpiness, but this is nit-picking. At an even more trivial level, the use of " = " for "equals" in paragraphs (in this book and A & K) is regrettable; but these imperfections cannot detract from the well-chosen examples, the pithy, enlightening commentary, and the elegant organization of this charming book.

Curtis employs an entirely different art form. Cayley and Sylvester, if they could return from their nineteenth century graves, might have trouble recognizing even their own work cast in the style of Bourbaki. Still, learning to read this style is important for an aspiring late-twentieth century mathematician, and the power of the approach is unquestionable.

In a book roughly the size of Smith's, Curtis covers essentially all of Smith's material and a great deal more. Among the new topics are polynomials over arbitrary fields, unitary (norm-preserving) transformations, duals of a vector space (linear transformations from the space to the real numbers), and the tensor product (a vector space $U \otimes V$ generated by any two original vector spaces U and V with an accompanying bilinear map t from $U \times V$ to $U \otimes V$ such that every bilinear map from $U \times V$ to every conceivable vector space W is the composition of t with a bilinear map from $U \otimes V$ to W). All of these steps occur on Curtis's road to the Spectral Theorem. But Curtis generalizes Smith's form of the theorem by omitting mention of matrices, using the complex field, and proving it for all normal operators, that is, all operators that commute with their adjoint—after defining "adjoint" by using dual spaces. (A self-adjoint operator is automatically normal since it is the same as its adjoint.) Curtis's Spectral Theorem asserts that any normal linear operator on a finite dimensional vector space can be written in the form

$$T = \lambda_1 E_1 + \lambda_2 E_2 + \cdots + \lambda_n E_n,$$

where the E_i are such that their sum is the identity transformation over the underlying vector space and their products are pairwise zero. It then follows that the E_i are idempotent ($E^2 = E$), but this is proved only incidentally, and nowhere in the book does Curtis use the graphic word "projection," a concept that would emphasize the fact that each normal operator can be viewed as a multiple of the identity individually on pairwise orthogonal invariant subspaces of the underlying vector space.

To his credit, however, he does warn the reader 90 pages earlier that "the search for invariant subspaces is the key to the deeper properties of a single linear transformation." Such considera-

tion of the reader's feelings is not reflected in his archaic use of "characteristic vector" for "eigenvector" and his nonstandard "diagonable." Such unnecessary trivialities are nuisances even to an experienced reader, and provide unnecessary obstacles to learning for graduates of a Matrix Theory course that used the standard words.

More serious are the implications of his recurring phrase (pages 1 and 163), "Everyone is familiar with..." followed by a concept which numerous people, many over four feet tall, do *not* know. Even respectable mathematics majors, when asked to recall some once familiar idea, have been known to squint their eyes, cock their heads, raise their shoulders as if to pounce on some elusive prey, and only finally nod. Statements suggesting that "everyone else" knows something that we don't can generate fleeting math anxiety even in experienced mathematicians. Curtis is not alone in his obliviousness to the majority of humankind, but it might help span the gulf between pure mathematicians and the rest of the world if he (and others) would express what he undoubtedly means in a more tolerant way. This is not to object to an occasional "clearly" when it saves printing costs and the reading audience is sufficiently advanced to know that "clearly" may abbreviate "It can be proved in less than two pages." But "It can be shown that..." is kinder to fledgling egos than "Clearly." Similarly, one might prefer Curtis to use, "It is assumed that the reader is familiar with... ."

Despite quibbles, Curtis provides a commendable exposition that links linear algebra to many other mathematical fields and provides a course more advanced than Smith's in both style and in content. However, despite the word "Introductory" in its title one can question whether Curtis' contents provide a sample undergraduate curriculum in these days of mass collegiate education. Surely there are brilliant undergraduates who would relish such a book, but the majority of Americans in college are still recovering from beginning algebra at the age of fourteen and are not likely to be prepared for Curtis before college graduation.

Springer-Verlag's caption "Undergraduate Texts in Mathematics" on the cover of its books makes it difficult to use them as graduate texts. It seems a shame that a great publisher does not make graceful allowances for the fact that colleges may need five or six years to prepare brilliant students from "disadvantaged" pre-collegiate environments for serious graduate work. Many colleges would find these books hopelessly advanced for undergraduates but tempting to share with Master's Degree candidates. It is humiliating for graduate students to use even a very fine text with "Undergraduate" blazoned on the cover, although this seems like an unfortunate reason to deselect a book as otherwise commendable as Curtis or Smith.

K & T was designed for a course somewhere between the above described Matrix Theory and Vector Space Theory courses. (A & T also serves this group if the entire book is used.) It aspires to take students who at first struggle to learn row reduction in the context of systems of linear equations through the mysteries of kernel and range to the application of eigenvectors to differential equations. It is a tall order for one semester. Consequently, it is difficult to originate an appropriate subtitle; "Compromise Linear Algebra" or "Vector Spaces Without the Spectral Theorem" would not look good in a college catalog.

K & T itself does not approach the sophistication of the spectral theorem, but it presents more of Differential Equations than most other linear algebra texts. One might wonder whether K & T could be the first text in a two-semester course following calculus that would cover both linear algebra and differential equations more thoroughly than two one-semester courses can do. This, of course, presents scheduling problems (and what would be the second semester text?), but K & T cleverly demonstrates how elementary differential equations can be used to motivate linear independence. It seems plausible, as the preface asserts, that students grow to understand linear independence in the process of finding complete solutions to second and higher order differential equations. In any case, K & T introduces them just before linear independence, and then says, "We shall explore the question of whether or not our spanning sets are as small as possible."

In diametric contrast to A & K, K & T postpones all mention of eigenvectors and eigenvalues until the end of the book. It also avoids matrix multiplication and inverses, hardly advanced topics, for its first 200 pages claiming that there is no motivation for them earlier. An elementary

course with applications can provide Leontief matrices or ecosystems to meet this need—but then we are back to a pre- or co-calculus course.

The difference between a "Linear Algebra: Matrix Theory" course similar to the first half of A & K (and who among us would claim that the second half of a text is always reached?) and a "Linear Algebra: Vector Space Theory" course like Curtis's may be even greater than the contrast between the Algebra 2 courses offered by a beleaguered inner city ghetto school and the most prestigious academy, a contrast exasperating to college faculties in many parts of the country. Just as the New Jersey Basic Skills requirement has given new hope to the undergraduate courses in New Jersey state colleges by enabling instructors to assume with some accuracy the preparation of the class, acknowledgement and accommodation for the disparity of undergraduate curricula would undoubtedly brighten the lives of university professors. Standardized tests are not only time-consuming to administer and impossible to judge fairly, but they do not measure the disciplined creativity that ultimately determines a student's success in graduate school.

However, a graduate department that genuinely wants as many as possible of its beginning students to bring honor to the department after earning degrees would be wise to send immediately a realistic, detailed summary of its prerequisite topics with suggested sources of such information to each student it admits. As a neophyte graduate student long ago, I longed for such a list. Lacking it, I compiled my own, got pregnant, read while the babies were sleeping, and returned with more adequate preparation—but this mode of coping is not available to most mathematics graduate students.

Similarly, each undergraduate institution that aspires to send some of its graduates to universities, even though it may be overworked and overworried serving the multitudes who are barely passing, should take time to post a sample of such lists for student perusal. It may be beyond an undergraduate department's fiscal and personnel capacities to provide good students with optimum instruction, but it would take little money or effort to thus inform them of a profitable way to spend their summer evenings.

No established American intellectual would clamor for the "top-down" curriculum standards common elsewhere. We know that a country without a central ministry of education can tailor its local offerings to the needs of individual students. However, if we are to continue to prepare the best students for viable competition in the international fray, it behooves us to acknowledge the unfortunate deception in subtitling a book like Curtis's "An Introductory Approach" and to give our students the information they need to thrive in a nation of superchoice.

An Introduction to Convex Polytopes. By Arne Brøndsted. Graduate Texts in Mathematics, vol. 90. Springer-Verlag, New York, 1983. vii + 160 pp.

CARL W. LEE
Department of Mathematics, University of Kentucky, Lexington, KY 40506

Geometric intuition and experience provide one with a general feeling of what a polytope is, at least in two or three dimensions. Nearly everyone has encountered the five Platonic solids at some time or another, and an acquaintance with the Archimedean solids is not unusual. The fascination with polytopes, and the desire to analyze and classify them, extend far back into history. Primary combinatorial questions usually concern the existence and/or enumeration of polytopes with certain numbers or types of faces of certain dimensions.

This book concerns itself with *convex polytopes*. We may formally define a convex polytope to be the convex hull of (i.e., the smallest convex set containing) a finite set of points in Euclidean space. Each of its *faces* is determined by an intersection with a "supporting" hyperplane meeting the polytope so that the polytope lies entirely within one of the two closed half spaces determined

by the hyperplane. The *dimension* of a face or of the polytope itself is given by the dimension of its affine span. The inclusion relation among the faces gives the set of faces of a polytope the structure of a lattice. Two polytopes have the same *combinatorial type* if they have isomorphic face lattices, and the central combinatorial problem is to characterize all of the different combinatorial types of convex d-dimensional polytopes.

If $f_j(P)$ denotes the number of j-dimensional faces of a polytope P, *Euler's relation* that $f_0(P) - f_1(P) + f_2(P) = 2$ for convex three-dimensional polytopes is the first fundamental combinatorial result in the theory of polytopes. Discovered by Euler in 1750, although strongly anticipated by Descartes more than one hundred years earlier, this theorem has deeply affected many branches of mathematics. Attempts to determine for which classes of objects Euler's relation holds threw into sharp focus the significance of convexity. Attempts to find analogous formulas for higher dimensional objects led to Schläfli's formulation of

$$f_0(P) - f_1(P) + \cdots + (-1)^{d-1} f_{d-1}(P) = 1 - (-1)^d$$

for d-dimensional convex polytopes. This formula is also referred to as *Euler's relation*. The first real proof of this d-dimensional version of Euler's relation was algebraic rather than strictly geometric, and was completed by Poincaré by now familiar techniques of algebraic topology. The quest for an entirely elementary geometric proof was ultimately successful. For example, in his beautiful and indispensable book on convex polytopes (*Convex Polytopes*, Wiley, 1967), Grünbaum offers such an argument, with no algebraic overtones.

In spite of Steinitz's characterization of face lattices of three-dimensional polytopes, the sheer difficulty of the central problem of classifying all of the combinatorial types of d-dimensional convex polytopes (d-polytopes) can be discouraging. One is therefore naturally inclined to investigate possibly easier subproblems. For example, one may restrict attention to *simplicial* polytopes (those for which every face is a simplex). For every convex polytope there is a (not necessarily geometrically unique) convex polytope having an anti-isomorphic face lattice (generalizing, for example, the duality exemplified between the octahedron and the cube); hence one may equivalently restrict attention to *simple* d-polytopes (those for which every vertex lies on exactly d $(d-1)$-dimensional faces. There are several advantages to this restriction. First, simplicial polytopes are a natural class in the sense that in a "general" finite set of points in $\mathbb{R}^d$ no more than d points lie on a common hyperplane. Second, if the feasible region of a nondegenerate linear program is bounded, then it is a simple polytope, and so many problems in mathematical programming correspond to problems on the combinatorial structure of simple polytopes. Third, a variety of problems about polytopes in general are reducible to ones concerning simplicial or simple polytopes. Finally, the face lattice of a simplicial polytope can be examined within the more general context of simplicial complexes, allowing the possibility of the application of certain algebraic techniques. Therefore, one is tempted to try to characterize all of the different combinatorial types of simplicial (or simple) convex polytopes.

A solution to this problem, however, also seems remote at present. Nevertheless, simplicial polytopes have proven to be more tractable than general ones. If we define the *f-vector* of a convex d-polytope P to be

$$f(P) = (f_0(P), f_1(P), \ldots, f_{d-1}(P)),$$

then Euler's relation specifies a linear relationship among the components of $f(P)$. This is the only linear relation that holds for all P, but for the subclass of simplicial polytopes there exists a set of exactly $[(d+1)/2]$ independent linear relationships, known as the *Dehn-Sommerville equations*. With such a strong set of necessary conditions, one might propose to characterize the set of all f-vectors of simplicial (or simple) convex d-polytopes.

A natural first step toward the solution of this problem is to determine minimum and maximum values of $f_j(P)$ given, for example, that P is a simplicial d-polytope with $n = f_0(P)$ vertices. Dually, one would want to find extremal values of $f_j(P)$ given that P is a simple

d-polytope with n $(d-1)$-dimensional faces. This was accomplished by the Lower Bound Theorem and the Upper Bound Theorem of Barnette and McMullen, respectively.

One of the main ingredients in McMullen's proof (see P. McMullen and G. C. Shephard, *Convex Polytopes and the Upper Bound Conjecture*, Cambridge University Press, 1971) was his use of a particular reformulation of the Dehn-Sommerville equations, which led him to a set of conditions that he conjectured would solve the characterization problem for f-vectors.

Then Stanley revealed the algebraic significance of this reformulation. He demonstrated that for a certain class of simplicial complexes, including boundary complexes of simplicial polytopes, there are naturally associated Cohen-Macaulay graded algebras. As a consequence of this, the Dehn-Sommerville equations turn out to be a manifestation of Poincaré duality. Stanley used this information to obtain a new proof of the Upper Bound Theorem and extended it to the strictly larger class of simplicial spheres as well. In the process, he determined new necessary conditions on the f-vectors of simplicial polytopes.

Ultimately, in 1979, Stanley made a further algebraic connection between simplicial convex polytopes and the cohomology of complex projective varieties that implied the necessity of McMullen's proposed conditions for f-vectors. At about the same time Billera and the reviewer succeeded in showing that McMullen's conditions were also sufficient to guarantee the realizability of an f-vector, completely solving the characterization problem. Whether the conditions also hold for simplicial spheres is an open question. Whether, as in the case of Euler's relation, an entirely elementary geometric proof of McMullen's conditions can be found, free of the elegant but elaborate techniques of algebraic geometry, also remains to be seen.

Brøndsted's book provides an introduction to the theory of convex polytopes, developing enough of the subject to be able to present proofs of Euler's relation, the Dehn-Sommerville equations, and the Lower and Upper Bound Theorems. Brøndsted chooses to demonstrate the latter three within the context of simple polytopes. Though not proving McMullen's characterization, he does show how the Lower and Upper Bound Theorems are corollaries of this more general result.

In the meantime, more information is being accumulated about the combinatorial structure of convex polytopes that may eventually lead to further breakthroughs in the central classification problems.

Mathematical People: *Profiles and Interviews*. By Donald J. Albers and G. L. Alexanderson, editors. Birkhäuser Boston, Cambridge, MA, 1985. vii + 372 pp. $21.00.

ANN HIBNER KOBLITZ
6547 17th Avenue N.E., Seattle, WA 98115

I'm a math groupie.* I confess it unabashedly. After all, there are many less pleasant groups of people with which one could fall into company than mathematicians. Speaking in generalities, I find that mathematicians, both historically and in the present, have many admirable qualities. They are tolerant of diversity (political as well as social and cultural), they are curious about the world, they have a quirky sense of humor and a keen feeling for the ridiculous, they tend to be elementally weird in ways I find endearing, even admirable.

For me, it comes as no surprise that the first woman in modern times appointed to a university

*For those of you not familiar with the term, or familiar only with the unfortunate sexual connotations of the word as used in the phrase "rock groupie," by groupie I mean a fan, an admirer, a partisan of mathematicians. In other words, when faced with groups of mathematicians, historians, social scientists, businessmen, etc., I prefer to hang around with the mathematicians.

chair in any field was a mathematician (Sofia Kovalevskaia), nor that when a survey of German professors was taken in 1896, mathematicians topped the list of those favoring the admission of women to universities (while historians were least favorable to women's higher education).

Historically, mathematicians have tended to take the attitude that anyone capable of proving good theorems should be welcomed into the mathematical community. This does not mean that they necessarily give any active encouragement to women, minorities, the economically disadvantaged, or the socially inept. Rather, this means that once talented members of the groups above have made their way through, around, and over the numerous obstacles society places in their way, mathematicians will more or less readily acknowledge their achievements and make a place for them in the mathematical world. (No, this does *not* automatically happen in other disciplines, contrary to what some engagingly naive mathematicians and many old-style Jacksonian democrats might think!)

As a consequence of this relatively tolerant attitude, one tends to find all sorts of people, including what an unsympathetic observer might characterize as an odd assortment of nuts and flakes, occupying respectable positions within the mathematical community. It is in part this aspect of the mathematical world that I find so fascinating, and I welcome any chance to increase my knowledge of the various types of people who work in mathematics and mathematics-related fields.

Mathematical People provides an excellent opportunity to acquaint oneself with the life stories and opinions of a large number of men and women whose livelihoods and avocations have something to do with mathematics. The majority of the selections in the book are interviews with famous and near-famous mathematical people, some of which first appeared in the *College Mathematics Journal*. The rest are profiles and autobiographical or biographical sketches; with the exception of the pieces on Raymond Smullyan and John Horton Conway, in my opinion these are the least successful segments in the collection. The book is sprinkled with portraits, drawings, photographs, all of which contribute to the pleasantly homey, casual tone sustained throughout most portions of *Mathematical People*.

On the whole, I found the collection fascinating. Most of the mathematical people interviewed expressed themselves well, and had intriguing, if sometimes idiosyncratic, opinions about virtually everything. To my mind, the most felicitous combinations of astute questions and informed, engaging responses were in the interviews with David Blackwell, Persi Diaconis, Martin Gardner, John Kemeny, Donald Knuth, Henry Pollack, Constance Reid and Herbert Robbins. Reading Blackwell and Diaconis on how they meandered their way into mathematical careers—or on any other topic either of them cared to tackle—is a delight. (Blackwell started out intending to become an elementary school teacher, while Diaconis left home at the age of fourteen for ten years on the road as a professional magician.)

A few of the selections are singularly infelicitous, either because the interview didn't gel, or because the text was insufficiently edited, or because the subject was essentially boring. (Yes, even among mathematicians one occasionally comes across a real yawner. But I only found one in *Mathematical People*, and he shall remain nameless.) And one or two comments by the elder statesmen of mathematics will make some readers squirm with vicarious embarrassment. In this connection, one might mention the charmingly childlike naïveté of Paul Halmos's statement on page 127: "I don't think mathematics needs to be supported. I think the phrase is almost offensive. Mathematics gets along fine, thank you, without money, and I look back with nostalgia to the good old days, three or four hundred years ago, when only those did mathematics who were willing to do it on their own time." Ah yes, bring back the good old days when only men of the gentry did mathematics!

I am not going to even try to be objective about *Mathematical People*. Everyone will have his or her own reaction to the book. Some people will no doubt throw it aside in disgust (an unwise move, however—*Mathematical People* is, if nothing else, a hefty projectile), while others will be indignant that I find any deficiencies in the collection at all. Some will be disappointed that their

favorite mathematical personalities were somehow missed (my missing favorites are Mark Kac, Serge Lang, and André Weil). But as the editors point out, that can't be helped. No collection of this type can hit everyone.

Mathematical People is not the type of book to be read cover to cover, or to be looked at when one is in a serious mood. Nor is it a book that will particularly appeal to historians of mathematics (at least not for the next fifty years or so) except as bedtime reading. And don't give it to Aunt Jane and Uncle Theodore as your last-ditch attempt to convince them that you and your colleagues are just like everyone else—it won't work.

On the other hand, if you're looking for something to read in front of the fire on a long winter evening, or want to show Aunt Jane and Uncle Theodore the quirky-but-charming side of your profession, or tend to give your immediate family books you know *you* will enjoy reading, then give *Mathematical People* a try. The collection is well constructed, and has something to interest and entertain anyone with any connection to the mathematical world.

A Course in Functional Analysis. By John B. Conway. Springer-Verlag, New York, 1985. xiv + 389 pp.

T. W. GAMELIN
Department of Mathematics, University of California at Los Angeles, Los Angeles, CA 90024

What is functional analysis? It is not easy to give a comprehensive defining statement. The boundaries of the field may be set in quite different ways, depending on how much noncommutative harmonic analysis or partial differential equations one embraces. At the core of functional analysis are normed spaces and more generally topological vector spaces, together with the functionals and operators thereon. The spaces usually come equipped with a potpourri of topologies. The hallmark of functional analysis is DUALITY, oftentimes in the form of the Hahn-Banach theorem in one of its manifold guises. The methods of duality fall under the rubric of "soft" analysis. When uttered by certain "hard" analysts, the word "soft" can assume a pejorative tone, yet even the most militant of the hard analysts do not blush upon invoking the Hahn-Banach theorem.

Functional analysis has served as a useful and unifying tool in numerous areas of mathematics, particularly differential equations and more recently stochastic theory. It has also become a fascinating and fruitful object of study for its own sake. The scope of the subject can be appreciated by reviewing its history, still fresh and quite interesting in its own right. The most complete and authoritative overview has been provided by Dieudonné in his *History of Functional Analysis*, reviewed by R. S. Doran for the *Bull. Amer. Math. Soc.*, 7(1982), pp. 403–409.

The field of functional analysis was virtually nonexistent before 1900. The seeds for the birth of functional analysis were sown towards the end of the nineteenth century, when certain problems stemming from the partial differential equations of mathematical physics were recast as integral equations of Fredholm type. The first decade of the twentieth century saw the birth and explosive growth of functional analysis, commencing with the seminal studies of Fredholm (1900–1903) of the class of integral equations which bears his name. Inspired by Fredholm, Hilbert (1906) investigated bilinear forms on infinite dimensional sequence spaces and established essentially the spectral theorem for compact symmetric operators. Meanwhile Lebesgue (1902) introduced his theory of integration, which was to provide the context for much of functional analysis. E. Fischer (1907) and Frigyes Riesz (1906) showed independently that the L^2-spaces associated with the Lebesgue integral are complete. By the end of the decade Riesz had introduced the L^p-spaces and obtained in essence the representation theorem for their duals. In the meantime, Fréchet had introduced in his thesis (1906) the fundamentals of metric spaces, paying special attention to the

space of continuous functions with distance defined by the supremum norm. It was a remarkable decade.

Among the various contributions to the nascent field of functional analysis, the work of Riesz stands out for its clarity and its modern tone. By 1913 Riesz had developed the spectral theory of bounded self-adjoint operators, including the functional calculus, and by 1916 he had developed his spectral theory of compact operators. Much of his work reads as if it were written only yesterday.

The decade of the 1920's saw the theory of normed spaces being developed by Banach and others, leading towards the appearance of Banach's classic monograph in 1932. The contraction mapping principle, a fundamental unifying tool in many areas of analysis and approximation theory, already appears in Banach's thesis (1920). The modern version of the Hahn-Banach Theorem was proved by Hahn in 1927. Weak topologies were defined and exploited.

At least one person was repelled by the explosive growth of the area. Norbert Wiener, who in 1920 independently discovered the axioms for Banach spaces, reports in his autobiography *I Am a Mathematician* that after contributing a few desultory papers he abandoned the theory in part on account of the "many inferior writers, hungry for easy doctor's theses, who were drawn to it."

The axiomatic foundations of Hilbert space theory did not appear until relatively late. They were laid down towards the end of the 1920's by von Neumann, who established the theory of unbounded operators on Hilbert space, and who proceeded in the 1930's in collaboration with Murray to found the theory of operator algebras (von Neumann algebras). In ground-breaking work accomplished over the years 1927–1935, Schauder extended Brouwer's fixed-point theorem to Banach spaces and combined this with certain "a priori" estimates to solve nonlinear partial differential equations. The theory of distributions was initiated by Sobolev (1936) and developed over succeeding years particularly by L. Schwartz. The work of Beurling in the 1930's served as a preamble to the theory of commutative Banach algebras, which was initiated by Gelfand in 1939.

The period of the 1950's and 1960's saw vigorous activity in a number of areas. The work of Grothendieck on topological tensor products and nuclear spaces was seminal. Pseudodifferential operators were developed as a tool for partial differential equations. The theory surrounding the Atiyah-Singer index theorem involved aspects of functional analysis, especially Fredholm operators. The techniques of commutative Banach algebras were initially applied to harmonic analysis, but later they were utilized by Bishop and Wermer in uniform approximation theory.

During the past decade or two there has been considerable activity in a number of areas within functional analysis. We mention some of these.

The theory of operator algebras, both von Neumann algebras and C^*-algebras, has seen intense activity. The current epoch of development of von Neumann algebras was signaled by the appearance of the Tomita-Takesaki theory aimed at classifying type III factors. The pivotal character in current activity is A. Connes, winner of the 1982 Fields Medal. Connes has made application of his non-commutative integration theory to foliations, and V. Jones has tied von Neumann algebras to knot theory. A recent development is Haagerup's proof of the uniqueness of hyperfinite type III_1 factors. On the other side of operator algebras, the C^*-algebras have provided a context for noncommutative topology. A signal result is the theorem of Brown, Douglas, and Fillmore, concerning extensions of C^*-algebras, which introduces K-theory into the subject.

In the realm of commutative Banach algebras, Dales and Esterle have independently solved in the negative the uniqueness-of-norm problem for $C(X)$. Their counterexamples depend on the axiom of choice. Solovay and Woodin have linked the problem to logic, showing that the answer is affirmative in some "reasonable" axiom systems. In recent years the most interesting developments in function algebras have been related to the notion of an analytic set-valued function, a unifying concept in spectral theory which was introduced by Oka in the 1930's but lay dormant for a number of years until it was recently rediscovered by Slodkowski.

The theory of Banach spaces has been rejuvenated and undergone extensive development in the past fifteen years, at the hands of Lindenstrauss, Pelczynski, and others. A number of classical problems have been settled. Enflo showed that there are separable Banach spaces without bases, and there are continuous linear operators without invariant subspaces. At the most recent of the international conclaves of Banach spacemen, held periodically at Kent State University, it seemed that the stage was dominated by J. Bourgain, whose difficult work is currently exerting a strong influence.

One of the signs of the times is the use of stochastic techniques in functional analysis. Functional analysis has always served as a useful tool in stochastic theory. Now martingales are not only permeating analysis, but they are finding their way into the geometry of Banach spaces and other aspects of functional analysis.

The author of a text on functional analysis has a rich history to reflect and a wide spectrum of topics from which to choose. After the heroic attempt of Dunford and Schwartz, and in deference to the pocketbook of the graduate student, no author has attempted to cover the field in fair detail. It is difficult enough to indicate the variform applications of functional analysis to other areas. In the case at hand, the author has chosen to limit his material with an eye towards his own research interests, which involve operators on Hilbert space.

The author assumes a solid background in integration theory, though he does provide in appendices material on the Riesz Representation Theorem for the dual of L^p and the Riesz-Kakutani Theorem representing the dual of $C(X)$. The first half of the book consists of an introduction to standard topics on normed spaces and linear operators, organized in such a way as to proceed from the special (Hilbert space in Chapters I and II) to the more general (Banach spaces in Chapter III, locally convex spaces and weak topologies in Chapters IV and V). The second half of the book is devoted to expositions of more specialized topics, usually in the framework of Hilbert space: Banach algebras (Chapter VII), C^*-algebras (Chapter VIII), normal operators (Chapter IX), unbounded operators (Chapter X), and Fredholm operators (Chapter XI).

The book includes an abundance of material, with a stated goal to "give as many examples as possible", yet one is left with the impression that the treatment is rather narrowly focused. A number of important topics are given short shrift or excluded. There is little on the geometry of Banach spaces. While the Krein-Milman theorem is covered, there is nothing of Choquet theory. Schauder's fixed-point theorem is proved and applied *à la* Lomonosov to polynomially compact operators. Otherwise there is very little that can be qualified as nonlinear functional analysis. The Fréchet derivative does not appear, nor does the contraction mapping principle. Early on Green's function is constructed for a second-order Sturm-Liouville operator and the associated integral operator is shown to be compact. Otherwise virtually no attention is paid to differential equations. Distributions are touched upon only briefly, in a section in which they are defined and shown to form the dual space of a locally convex space. The Gelfand-Naimark-Segal construction is presented, but irreducible representations are not discussed. There is no mention of factors, nor of extensions of C^*-algebras, nor of K-theory. While Stone's theorem for strongly continuous unitary groups is proved (Chapter X), there is no coverage of semigroups of operators, much less the differential equations which give rise to them. Historical notes are not included, and no sense of the origins or history of functional analysis is imparted. My own preference would be for more ideas for applications, for more of the excitement of current research directions, for more loose ends, at the expense of attention to the more formal development of the theory.

The book does provide a substantial introduction for the student who is preparing to do research in operator theory. The treatment of normal operators on Hilbert space is particularly thorough. Some theorems are stated without proof, and these whet the reader's appetite for more. The book is well written, it is easy to read, and it is easy to learn from. The author has an engaging style which is a good blend of the formal and the informal. In spite of its restricted focus, the book will provide a number of students of mathematics with a readable and enjoyable source for functional analysis, particularly as it relates to operator theory.

MACSYMA gives you the answers.

```
(C1) DEPENDS(Y,T)$
(D1)                    [Y(T)]
(C2) DIFF(Y,T)+Y^ 2+(2*T+1)*Y+T^ 2+T+1;
```

$$\text{(D2)}\qquad \frac{dY}{dT} + Y^2 + (2T+1)Y + T^2 + T + 1$$

```
(C3) SOLN:ODE(%,Y,T);
```

$$\text{(D3)}\qquad Y = -\frac{\%C\ T\ \%E^T - T - 1}{\%C\ \%E^T - 1}$$

```
(C4) SUBST([T = 1,Y = 1], %);
```

$$\text{(D4)}\qquad 1 = -\frac{\%E\ \%C - 2}{\%E\ \%C - 1}$$

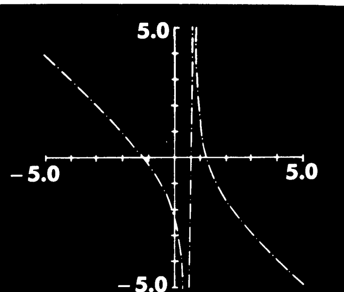
```
(C5) SOLVE(%,%C),NUMER;
(D5)                  [%C = 0.5518192]
(C6) SPECIFIC_SOLN:SUBST(%, SOLN);
```

$$\text{(D6)}\qquad Y = -\frac{0.5518192\,T\ \%E^T - T - 1}{0.5518192\ \%E^T - 1}$$

In symbolic form.

```
(C7)   FORTRAN(%)$
          Y = -(0.5518192*T*EXP(T)-T-1)
               /(0.5518192*EXP(T)-1)
```

In FORTRAN code.



$$Y = -\frac{0.5518192T\ \ \%E^T\ - T\ - 1}{0.5518192\ \ \%E^T\ - 1}$$

In graphics.

In seconds.

Put MACSYMA's Artificial Intelligence to work for you.

If it is still taking you weeks or months to handle rudimentary algebraic problems and extremely complex analyses or to generate FORTRAN code out of symbolic expressions, then make it your business to find out about MACSYMA mathematical software. It is available right now from Symbolics, Inc., the leader in symbolic processing. For a literature kit including a full capabilities brochure, just clip and mail the coupon or **call toll-free.**

1-800-MACSYMA

In Mass., Alaska & Hawaii only call **(617) 577-7500.**

MACSYMA™

The most comprehensive software approach to symbolic mathematical computations.

symbolics™

Your next step in computing.™

Symbolics, Inc. Dept. M-AM1
11 Cambridge Center
Cambridge, MA 02142

Name ______

Title ______

Company ______

Address ______

City ______ State ______ Zip ______

Telephone ______ Extension ______

MACSYMA is available to colleges and universities at special rates.

The Mathematical Association of America
CORPORATE MEMBERS

The Mathematical Association of America gratefully acknowledges the interest and support of the following:

CORPORATE MEMBERS

Addison-Wesley Publishing Company, Inc.
American Mathematical Society, Inc.
Birkhäuser Boston, Inc.
Digital Equipment Corporation
International Business Machines Corporation
Macmillan Publishing Company
Marcel Dekker, Inc.
The Mitre Corporation
Pergamon Press, Inc.
Prindle Weber and Schmidt, Inc.
Saunders College Publishing, Inc.
Scott Foresman & Company, Inc.
Springer-Verlag New York, Inc.
John Wiley & Sons, Inc.

CONTRIBUTING CORPORATE MEMBER
Bell Laboratories, Inc.

for your library

The William Lowell Putnam Mathematical Competition, 1965-1984,
compiled by Gerald L. Alexanderson,
Leonard Klosinski and Loren Larson
151 pp., Cloth, ISBN 0-88385-441-4
List: $24.00 MAA Member: $18.00

The Putnam Competition has since 1928 been providing a challenge to the gifted college mathematics student. This volume contains problems with their solutions, for the years 1965-1984. Corrections to the solutions have been made, and additional solutions are presented. Included is an essay on recollections of the first Putnam Exam by Herbert Robbins, as well as appendices listing the winning teams and students from 1965 through 1984.

The earlier collection of Putnam problems, **The William Lowell Putnam Mathematical Competition, 1938-1964** was a scholarly approach to the problems which gave many different solutions for each problem. The authors gave us a background, in depth, on each problem and showed us, whenever possible, how the problems stimulated further questions. The current volume of Putnam problems does not attempt to duplicate that effort, but it does offer the problem solver an enticing sample of challenging problems and their solutions.

Order from:
The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036

INDEX TO VOLUME 93, 1986
THE AMERICAN MATHEMATICAL MONTHLY

AUTHOR INDEX

ERRATA AND ADDENDA

H. William Oliver has pointed out an oversight in "An elementary proof of a theorem in calculus" (D. E. Richmond, this MONTHLY, 92 (1985) 589). The author argues that since $[u_n, v_n]$ converges to x and

$$\frac{f(u_n) - f(v_n)}{u_n - v_n} \geq C \geq 0 \quad \text{and} \quad f'(x) = 0,$$

then necessarily $C = 0$. Oliver notes that the validity of this assertion depends upon x being between u_n and v_n, in which case

$$C \leq \frac{f(u_n) - f(x)}{u_n - x} + \frac{f(x) - f(v_n)}{x - v_n}$$

and Richmond's conclusion is obtained.

L. E. Mattics and F. W. Dodd state that they have been able to improve their result $f(n) \leqslant n$ in their paper "A bound for the number of multiplicative partitions", this MONTHLY, 93 (1986) 125–126, to $f(n) \leqslant n/\log(n)$ for all $n > 1$, $n \neq 144$, thus settling a second conjecture of Hughes and Shallit. They also wish to inform the readers of the article, "On a problem of Oppenheim concerning 'Factorisatio Numerorum'," by Canfield, Erdős, and Pomerance (Journal of Number Theory, 17 (1983) 1–28) which contains some remarkable asymptotic results for $f(n)$.

SUBJECT INDEX

This index uses the current version of AMS 1980 Mathematics Subject Classification.

MISCELLANEA

INDEX OF SPECIAL RECURRING TOPICS

REBUSES

PROBLEMS AND SOLUTIONS

PROBLEMS PROPOSED

PROBLEMS SOLVED

SOLUTIONS

Numbers in boldface refer to problems; those in lightface to pages.

REVIEWS

Names of authors are in ordinary type; those of reviewers in capitals.

LETTERS TO THE EDITOR

PHOTOS

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 93, Number 10 December 1986

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. (This is necessary for indexing purposes.) Send all main articles, letters to the editor, and editorial correspondence to the editor-elect: HERBERT S. WILF, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Contributions to the other four special departments (Notes, Teaching of Mathematics, Unsolved Problems, and Problems and Solutions) should be sent to the appropriate one of the associate editors, as follows:

Notes: CAROL G. CRAWFORD, Department of Mathematics, U.S. Naval Academy, Annapolis, MD 21402, *or* RICHARD LIBERA, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716, *or* ANITA E. SOLOW, Department of Mathematics, Grinnell College, Grinnell, IA 50112.

Teaching of Mathematics: JOAN P. HUTCHINSON *or* STAN WAGON, Department of Mathematics, Smith College, Northampton, MA 01063.

Unsolved Problems: RICHARD GUY, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.

Proposed problems and solutions, both elementary and advanced: PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801.

Reprint Permission: A. B. WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Back Issues: P. and H. BLISS Co., Middletown, CT 06457.

Microfilm Editions: University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

Editorial Assistants: Fanny Hailpern, Mary P. Jackson.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT, ten times a year: January, February, March, April, May, June-July, August-September, October, November and December.

The annual subscription price for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is $24 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $24. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $70 per year.

Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

PRINTED IN THE UNITED STATES OF AMERICA

CRYSTALLOGRAPHY AND COHOMOLOGY OF GROUPS

HOWARD HILLER
Department of Mathematics, Columbia University, New York, NY 10027

Introduction. One imagines a crystal as an interlocking system of molecules (or balls connected by rods) that can be continued indefinitely in any direction filling up all of space (see Fig. 1). The essential feature that a mathematical treatment of crystals should abstract is the existence of a "small chunk" of the crystal pattern which acts as a building block for the whole structure. Of course, this building block must be of a very special nature so that the pieces will fit together.

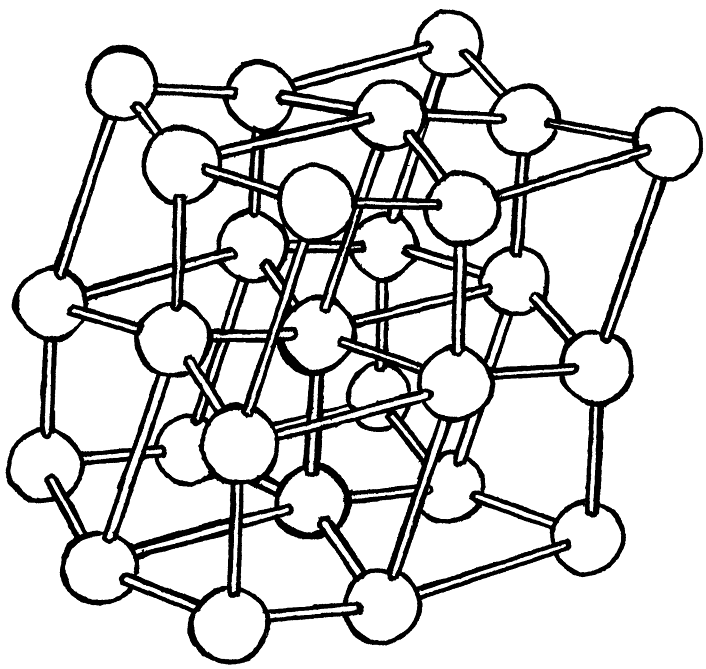


FIG. 1

The mathematical approach to this problem is to replace the crystal pattern by the group G of rigid motions of Euclidean space $\mathbb{E}^n$ that preserve it. Such a group is called a **space group**. In this group G there are n linearly independent translations that correspond to the existence of the building block described above. These translations generate a subgroup M of G which is free abelian (i.e., is isomorphic to the sum of n copies of the group $\mathbb{Z}$ of integers) and is called the **lattice** of G.

But there are also other symmetries the crystal system might possess (for example, rotations). If we consider the quotient group G/M, we obtain another algebraic invariant of the space group G called the **point group** H.

Unfortunately these two invariants: the lattice M and the point group H are not enough to determine the space group G. We give an example in dimension 2, where a crystal pattern

Howard Hiller was born in Brooklyn, New York, and educated at Cornell University and MIT. His research interests have varied from mathematical logic to algebraic K-theory to differential geometry. He has spent a year at Oxford University under an American Mathematical Society Research Fellowship, a year at Göttingen under an Alexander von Humboldt Fellowship, and taught at Yale University and Columbia University. Currently he is an assistant vice-president at Citicorp Investment Bank.

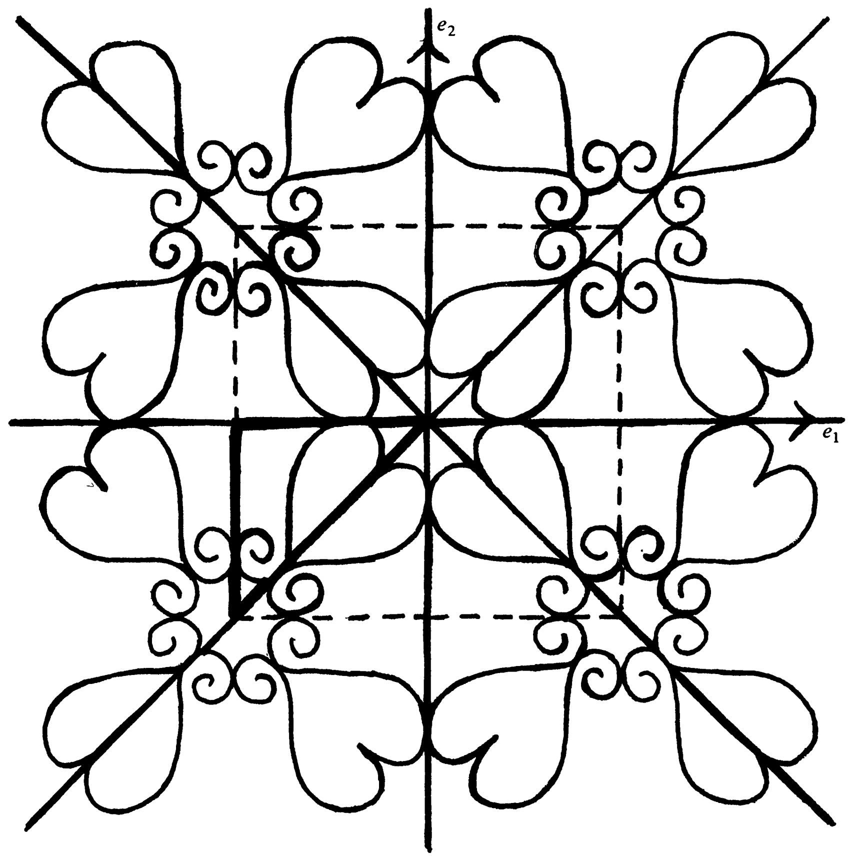


FIG. 2(a) G. The bold lines indicate reflections that are in the symmetry group. The boldly outlined triangle near the center is the fundamental domain for the symmetry group. The fundamental domain for the lattice generated by the translations in the directions e_1 and e_2 is the central dashed square. Rotation of 90° around the center is also in the symmetry group.

corresponds to a "tiling" of the plane. Consider the two tilings in Fig. 2.

Both symmetry groups G, G' contain translations in the directions $\mathbf{e}_1$, $\mathbf{e}_2$, the standard basis vectors. It can also be shown that in both cases the point group H is D_4, the symmetry group of a square, of order 8. (What are the 8 rigid motions that preserve a square ?) But these two tilings are not the same! The bold lines in the figures represent reflection mirrors. The group G has a center of a 90° rotation that lies on the intersection of two reflection mirrors. The group G' lacks such a center. Hence they cannot be the same.

The goal of this survey is to try to find the missing algebraic invariant that will determine the space group and then use this invariant to classify and enumerate space groups. The invariant is found in the cohomology of groups (see Section 5). This invariant plays a role in the modern solution to the eighteenth problem of Hilbert (see Section 7) namely a proof that in each dimension n there are only finitely many space groups. Furthermore, the proof can be turned into an algorithm for enumerating space groups and hence gives a dimension-independent approach to the crystallographer's enumeration of the 219 crystals in dimension three.

The prerequisites for reading this article have been kept as minimal as possible. Nonetheless the reader should have seen some topological notions as in a course on advanced calculus, have a working knowledge of linear algebra, and some familiarity with basic group theory.

1. Euclidean geometry. We let $\mathbb{R}^n$ denote a real vector space of dimension n. A point in $\mathbb{R}^n$ is

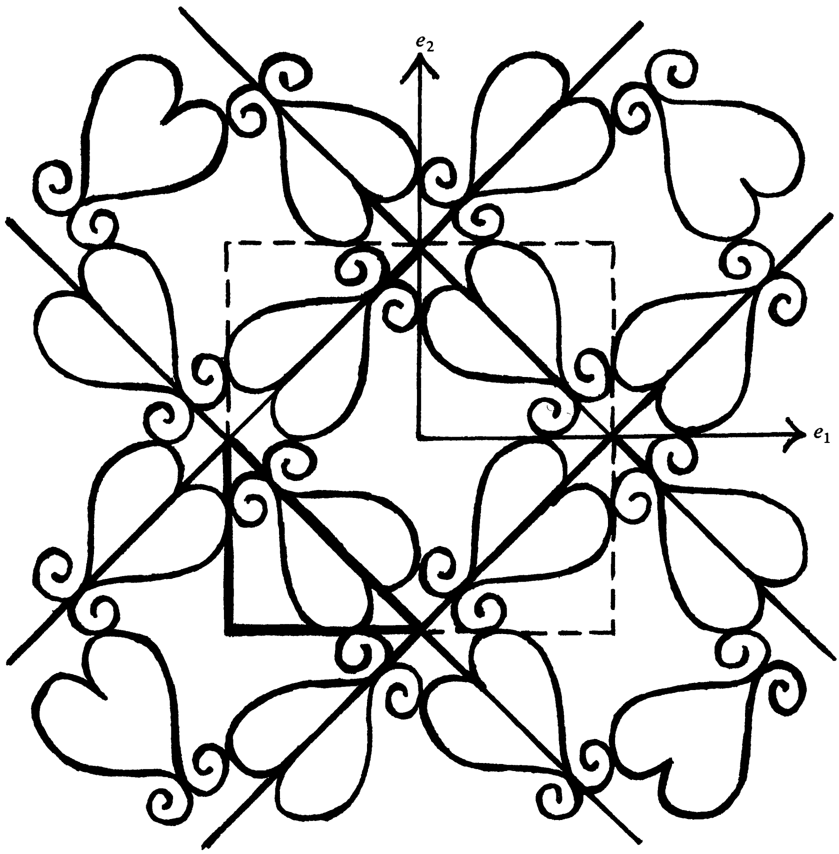


FIG. 2(b) G′

specified by an n-tuple $(x_1, \ldots, x_n)$ of real numbers. These points are added component-wise and can be multiplied component-wise by real numbers.

In order to do geometry we must equip this space with additional notions of length and angle. This can be done efficiently with the notion of the **inner** (or **dot**) product of two vectors. If x and y are in $\mathbb{R}^n$, we define the inner product

$$x \cdot y = \sum_{i=1}^{n} x_i y_i,$$

where $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$. Now we can define the **length** of a vector x as

$$\|x\| = (x \cdot x)^{1/2},$$

the **angle** between two non-zero vectors x, y by

$$\theta = \arccos(x \cdot y/\|x\|\,\|y\|),$$

and the **distance** between x and y by $\|x - y\|$.

We refer to our space equipped with this additional structure as Euclidean space and denote it by $\mathbb{E}^n$. An **isometry** (or **rigid motion**) of $\mathbb{E}^n$ is a mapping $f\colon \mathbb{E}^n \to \mathbb{E}^n$ that preserves distance, i.e., $\|f(x) - f(y)\| = \|x - y\|$, for all x, y in $\mathbb{E}^n$. It is not difficult to show that f is necessarily bijective (see [**21**, Chap. 3]).

A good example of an isometry is an **orthogonal** mapping. This is an invertible linear mapping that preserves the inner product and, in particular, fixes the origin. The set of such mappings

forms a subgroup $O(n)$ of the group of all real invertible linear mappings called the **orthogonal group**. If we identify a linear mapping f with the matrix A_f that represents it (say, with respect to the standard basis), then the orthogonality condition can be written $AA^t = I$, where $(\ )^t$ denotes the transpose of a matrix and I is the identity matrix.

If, in addition, the linear mapping has determinant 1, we get a smaller subgroup written $\mathrm{SO}(n)$ and called the **special orthogonal group**. For example, if we choose a line (an axis) in $\mathbb{R}^n$ and rotate about it, we get an element of $\mathrm{SO}(n)$. If we choose a hyperplane in $\mathbb{R}^n$ and reflect through it, we get an element of $O(n)$ of determinant -1 (hence not in $\mathrm{SO}(n)$).

If we fix a vector v in $\mathbb{E}^n$, the translation mapping $t(v)$: $\mathbb{E}^n \to \mathbb{E}^n$ that sends a vector x to $x + v$ is also an isometry of $\mathbb{E}^n$. We write $V = \{t(v)\colon v \in V]$ for the vector space of translations (ignoring the inner product structure now) to distinguish it from the Euclidean space $\mathbb{E}^n$. Note that a translation determined by a nonzero vector is an example of an isometry that is not a linear mapping (as the origin is not even fixed).

The main result of Euclidean geometry (see [**21**, p. 101]) is the following assertion:

(1.0) THEOREM. *Every isometry can be written in a unique way as a composition $t(v) \circ \phi$, where $t(v)$ is the translation by v and ϕ is an orthogonal mapping of $\mathbb{E}^n$.*

We can express this result very neatly in the language of group theory. Let us write $\mathrm{Isom}(\mathbb{E}^n)$ for the set of all isometries of $\mathbb{E}^n$. It is not difficult to check that under the operation of composition of mappings $\mathrm{Isom}(\mathbb{E}^n)$ is a group. It is called the **Euclidean group** or the **group of rigid motions** of $\mathbb{E}^n$. It is not difficult to check (see (1.1iii) below) that the vector space V of translations is a normal subgroup of $\mathrm{Isom}(\mathbb{E}^n)$. It is a consequence of Theorem (1.0) that the quotient is isomorphic to the orthogonal group $O(n)$. This allows us to describe the elements of the isometry group and their composition in a very concrete fashion. Let (v, ϕ) denote the element $t(v) \circ \phi$ of $\mathrm{Isom}(\mathbb{E}^n)$ and note that the action of this element (v, ϕ) on a vector x in $\mathbb{E}^n$ is given by

$$(v, \phi)(x) = t(v)(\phi(x)) = v + \phi(x).$$

Hence it is easy to figure out what the multiplication in $\mathbb{E}^n$ must be. We compute

$$\begin{aligned}(v, \phi)(v', \phi')(x) &= (v, \phi)(v' + \phi'(x)) \\ &= v + \phi(v' + \phi'(x)) \\ &= (v + \phi(v'), \phi\phi')(x).\end{aligned}$$

This forces that we define the multiplication in $\mathrm{Isom}(\mathbb{E}^n)$ by

$$(v, \phi)(v,', \phi') = (v + \phi(v'), \phi\phi').$$

It is now easy to check the following assertion:

(1.1) LEMMA. (i) *The identity element of* $\mathrm{Isom}(\mathbb{E}^n)$ *is* $(0,1)$.

(ii) *The inverse of an element of* $\mathrm{Isom}(\mathbb{E}^n)$ *is given by*

$$(v, \phi)^{-1} = (-\phi^{-1}(v), \phi^{-1}).$$

(iii) *The action of* $\mathrm{Isom}(\mathbb{E}^n)$ *on its subgroup V is given by*

$$(v, \phi)(t,1)(v, \phi)^{-1} = (\phi(t),1).$$

Proof. The first part is completely trivial and for the second we have

$$(v, \phi)(-\phi^{-1}(v), \phi^{-1}) = (v + \phi(-\phi^{-1}(v)), \phi\phi^{-1}) = (0,1).$$

Finally, (iii) follows from

$$(v, \phi)(t,1)(v, \phi)^{-1} = (v + \phi(t), \phi)(-\phi^{-1}(v), \phi^{-1})$$

$$= \left(v + \phi(t) + \phi\left(-\phi^{-1}(v)\right), \phi\phi^{-1}\right)$$
$$= (\phi(t), 1).$$

Notice that the multiplication on Isom ($\mathbb{E}^n$) is **not** the usual direct product group structure. In the next section we develop the group theory that will clarify this new type of "product".

2. Extensions of groups. If K and H are two groups we can define a group structure on the set-theoretic product $K \times H$ by defining a multiplication pointwise. Namely

$$(k, h) \cdot (k', h') = (kk', hh').$$

This is the **direct product** of K and H, which we write $K \times H$. By an **action** of H on K we mean a group homomorphism α: $H \to \mathrm{Aut}(K)$. In particular, if h is in H and k is in K, we can describe the action by

$$h \cdot k = \alpha(h)(k).$$

With such an action we can define the **semi-direct product** group structure $K \rtimes_\alpha H$ on the set-theoretic product $K \times H$ by the multiplication.

$$(k, h) \cdot_\alpha (k', h') = (k + \alpha(h)(k'), hh').$$

This may seem on the face of it to be a highly unmotivated construction, but if we compare it with our geometric situation it is perfectly natural. If we let $K = V$, $H = O(n)$ and let α: $O(n) \to \mathrm{Aut}(V)$ be the natural inclusion then we can conclude from Section 1

$$\mathrm{Isom}(\mathbb{E}^n) = V \rtimes_\alpha O(n).$$

In fact, it is possible to give a more abstract characterization of semi-direct products. If K, G and H are groups, K abelian, we say that the following diagram of groups and homomorphisms

$$K \xrightarrow{i} G \xrightarrow{p} H$$

is **exact** if the map i includes K as a normal subgroup of G and p is a surjection of G onto H that induces an isomorphism of groups $G/i(K) \cong H$. One also says that G is an **extension** of K by H.

There is an induced action of H on K by pulling-back and conjugating. More precisely, if k is in K, then $h \cdot k = h^{\sim} k(h^{\sim})^{-1}$ in K, where $h^{\sim}$ is any element of G satisfying $p(h^{\sim}) = h$. This definition is independent of the choice of $h^{\sim} \in G$, as the reader should check (K is abelian). Hence in this situation it makes sense to ask if G is the semi-direct product of K by H via this action. The answer is yes if and only if there exist a **splitting homomorphism** σ: $H \to G$, i.e., a group homomorphism for which $p \circ \sigma$ is the identity map on H. (There is always a set-theoretic map satisfying this condition but we are insisting that it be a homomorphism of groups.) Clearly such a map exists if $G = K \rtimes_\alpha H$ by taking $\sigma(h) = (1, h)$, where 1 is the identity element of K. We often say the sequence is **split** by σ.

Another example of a semi-direct product is the **affine group** Aff ($\mathbb{E}^n$) of Euclidean space $\mathbb{E}^n$. It can be written $\mathrm{Aff}(\mathbb{E}^n) = V \rtimes \mathrm{GL}(n, \mathbb{R})$, where $\mathrm{GL}(n, \mathbb{R})$ is the group of invertible $n \times n$ matrices with real entries. These affine mappings need not preserve distance.

The ideas from this section will play a crucial role in our algebraic understanding and classification of crystals.

3. Space Groups. The fundamental notion that relates the group theory of Section 2 with the geometric ideas of crystallography is the notion of a fundamental domain. This requires a certain familiarity with topological notions. If a group G acts on a subset X of $\mathbb{R}^n$, a **fundamental domain** for this action is an open subset D of X that satisfies two properties:

(a) $\cup\{\text{closure } (gD) \colon g \in G\} = X$,
(b) $D \cap gD = \varnothing$, for all $g \neq 1$ in G.

Consider $X = \mathbb{E}^2$ and $G = \mathbb{Z} \times \mathbb{Z} \subset V \subset \operatorname{Isom}(\mathbb{E}^2)$ acting on $\mathbb{E}^2$ as a group of translations. The open unit square $D = (0,1) \times (0,1)$ is a fundamental domain for this action as the reader can easily check.

Another way to view this is by considering the quotient of the action of G on X. The action of G on X defines an equivalence relation $\equiv_G$ on G by $x \equiv_G y$ if there exists a g in G satisfying $g \cdot x = y$. There is a (topological) quotient space $X/G = X/\equiv_G$. In the example described above the quotient $\mathbb{E}^2/\mathbb{Z} \times \mathbb{Z}$ is a torus. The vertical translation produces a cylinder and the horizontal translation identifies the two "ends" of this cylinder to produce a torus. The torus is compact. We come now to the fundamental definition:

(3.1) DEFINITION. A discrete subgroup G of $\operatorname{Isom}(\mathbb{E}^n)$ is a **space group** (or **crystallographic group**) if the quotient space $\mathbb{E}^n/G$ is compact.

The reader should convince himself that this condition is equivalent to the compactness of the closure of the fundamental domain of G acting on $\mathbb{E}^n$. The discreteness condition means that if x_0 is in $\mathbb{E}^n$ the set $\{g \cdot x_0 \colon g \in G\}$ has no accumulation point.

The study of such space groups was motivated in part by Hilbert's eighteenth problem (see section 7) and led Bieberbach to the following characterization.

(3.2) BIEBERBACH'S FIRST THEOREM. *A subgroup G of* $\operatorname{Isom}(\mathbb{E}^n)$ *is a space group if and only if G contains n linearly independent translations.*

The crucial and difficult part of this result is the "only if" direction and a modern account of it can be found in Wolf's book [**30**]. A new and more informative proof has been recently discovered by M. Gromov. An exposition of his work can be found in [**7**].

The characterization of space groups provided by Bieberbach's First Theorem is not completely satisfying. It depends in an essential way on the realization of the elements of the group as isometries of $\mathbb{E}^n$. It would be preferable to have a purely algebraic (i.e., intrinsic) characterization of the class of space groups independent of their embedding inside $\operatorname{Isom}(\mathbb{E}^n)$. We have already made some progress in this direction by showing that G fits into an exact sequence

$$M \to G \to H,$$

where M denotes the free abelian group (isomorphic to the direct sum of n copies of the group $\mathbb{Z}$ of integers) generated by the translations provided by Bieberbach's First Theorem. This group M is often called the **lattice** of G and $H = G/M$ is called the **point group** of G. It is possible (although not trivial) to show that the quotient H is finite (see [**26**, pp. 26-27]). In fact, more is true. The group H acts on M by pulling an element h in H back to $h^{\sim}$ in G and conjugating as in Section 2. One can show that this action is faithful (see [**26**, p. 30]). Finally we have the following assertion:

(3.3) THEOREM (Zassenhaus [**31**]). *An abstract group G is isomorphic to an n-dimensional space group if and only if G contains a finite index, normal, free abelian subgroup of rank n, that is also maximal abelian.*

The maximal abelian property is a direct reformulation of the faithfulness of the above action of H on M. (Check this.) This Theorem provides the desired purely group-theoretic characterization of a space group. The problem remains to determine some practical set of algebraic data that specifies G and can be used to enumerate space groups. We return to this problem in Section 4.

Now that we have a notion of space group, we must decide when we want to consider two such to be equivalent. One possibility is to consider two space groups identical if they are abstractly isomorphic as groups. This is apparently the point of view taken by Bieberbach [**10**] in his landmark paper on the subject in 1910. A year later Frobenius [**11**] suggested that perhaps a more intrinsic notion would be preferable. He considered two space groups to be equivalent if when viewed as subgroups of the Euclidean group $\operatorname{Isom}(\mathbb{E}^n)$ they are conjugate by an element of the somewhat larger affine group $\operatorname{Aff}(\mathbb{E}^n)$. This more geometric notion is called **affine equivalence**.

Soon after, Bieberbach published his second fundamental paper on the subject [**4**] and in fact showed the following result:

(3.4) BIEBERBACH'S SECOND THEOREM. *Any abstract isomorphism of space groups can be realized by conjugation by an affine motion of* $\mathbb{E}^n$.

Hence Bieberbach and Frobenius were using the same equivalence relation anyway. This basic fact (often referred to as "rigidity") will also turn up again in Section 5.

4. Crystal classes. Now that we feel comfortable deciding when two space groups should be considered the same, we can try to group them into naturally defined classes. The most obvious parameters for classifying space groups are the ones we have already introduced–namely the point group H, the lattice M, and the action of H on M. For convenience we call these three pieces of structure together a **crystal class** and denote it simply (H, M). The reader is warned that the notation is a bit sloppy in that mention of the action itself is suppressed.

The study of actions of finite groups on lattices ($\cong \mathbb{Z}^n$) is a rich and important subject in itself called **integral representation theory**. The standard reference is [**8**]. Sometimes the techniques and results of integral representation theory can be profitably brought to bear on problems in crystallography.

As in the previous section we are faced with the problem of deciding when to consider two crystal classes "equivalent". In contrast with our treatment of space groups we find that there are two reasonable notions of "equivalence" that are clearly not identical. Both of them will be useful in our treatment of the classification of space groups.

If we choose a free integral basis for M, a crystal class can be viewed as a one-to-one homomorphism $f\colon H \to \mathrm{Aut}(M) \cong \mathrm{GL}(n, \mathbb{Z})$, where $\mathrm{GL}(n,\mathbb{Z})$ denotes the general linear group of non-singular $n \times n$ integer entries and determinant ± 1. (Recall from linear algebra that this last condition is equivalent to the condition that all the entries of the inverse matrix are integers.) So after choosing a basis for M, H is embedded explicitly as a subgroup $f(H)$ of $\mathrm{GL}(n,\mathbb{Z})$. If $f'\colon H \to \mathrm{Aut}(M')$ is another crystal class, we say (H, M) is **arithmetically equivalent** (or **$\mathbb{Z}$-equivalent**) to (H, M') if there exists an isomorphism $\alpha\colon M \to M'$ so that

$$\alpha \cdot f(h) = f'(h) \cdot \alpha$$

for all h in H. If we write this condition as $\alpha \circ f \circ \alpha^{-1} = f'$, we see that this is equivalent to insisting that the two corresponding subgroups $f(H), f'(H)$ are conjugate in $\mathrm{GL}(n,\mathbb{Z})$. If the two subgroups are conjugate in the larger rational general linear $\mathrm{GL}(n,\mathbb{Q})$ (where the matrix entries are rational numbers and the determinant is a nonzero rational number), then we say that the two crystal classes are **geometrically equivalent** (or **$\mathbb{Q}$-equivalent**). The resulting equivalence classes are the **arithmetic** and **geometric crystal classes**, respectively.

As an example consider the following three matrices in $\mathrm{GL}(n,\mathbb{Z})$:

$$A_1 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Since each has the property that $A^2 = 1$, they determine three crystal classes of the group $H = \mathbb{Z}_2$, the integers modulo 2. The matrix A_1 has eigenvalues $-1, -1$ and both A_2 and A_3 have eigenvalues $1, -1$ (compute it for A_3). Hence A_1 determines a different geometric crystal class than A_2, A_3. On the other hand we also have

$$\alpha A_2 \alpha^{-1} = A_3 \quad \text{where} \quad \alpha = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

So A_2, A_3 determine the same geometric crystal class. But since $\det(\alpha) = 2$, α is not in $\mathrm{GL}(2,\mathbb{Z})$ and α does **not** show that A_2, A_3 determine the same arithmetic crystal classes. We leave it as an exercise to show that A_2 and A_3, in fact, determine different arithmetic crystal classes.

We see from these samples that a single geometric crystal class can break up into two different arithmetic crystals classes. (Another example in dimension 2 is the group $H = D_3$, the symmetry

group of the triangle with 6 elements; see Section 8 below). It is a consequence of a fundamental result of integral representation theory that a geometric crystal class can split up into at most finitely many arithmetic crystal classes. This is the Jordan-Zassenhaus theorem [**8**, Theorem 79.1] and is a key ingredient in the solution of Hilbert's eighteenth problem.

5. Cohomology. If G is a space group, G fits into a short exact sequence

$$(*) \qquad M \xrightarrow{i} G \xrightarrow{p} H,$$

where M is free abelian, H is finite and H acts faithfully on M, so determines an arithmetic crystal class (H, M) (again our notation is somewhat sloppy). We will also say that G is **in the arithmetic crystal class** (H, M). Although G can be thought of as sitting inside the semi-direct product $V \rtimes_\alpha O(n) \cong \mathrm{Isom}(\mathbb{E}^n)$, (as in Section 2) there is no *a priori* reason to believe that G itself is the semi-direct product of H acting on M. In fact, this is the heart of our method for classifying space groups, seeing how far they differ from the semi-direct product. (The crystallographers call space groups that are semi-direct products **symmorphic**). We will see that Fig. 2 is a concrete geometric example of this phenomenon in dimension two.

By Bieberbach's First Theorem we can suppose that $G \subset V \rtimes_\alpha O(n)$ and that the map p: $G \to H$ is projection onto the second factor. Suppose τ: $H \to G$ is a set-theoretic section to this map p: $G \to H$. This means that for each element h in H, $p(\tau(h)) = h$. So $\tau(h) = (\sigma(h), h)$ for some set-theoretic map σ: $H \to V$. We will refer to this map σ as a **section** to the exact sequence $(*)$ and dispense with τ altogether.

If h happens to actually be in G, i.e., really is a symmetry of the crystal, then $\sigma(h)$ can be chosen to be any element of M, for example 0. But there are, for example, glide reflections (a reflection followed by a translation), for which $\sigma(h)$ cannot be chosen to be in M. We also know from Section 2 that if τ can be chosen to be a group homomorphism, then G is necessarily the semi-direct product of H acting on M.

Of course, such a section σ is not unique. We can easily remedy this situation by composing with the natural projection $V \to V/M$. The resulting map s: $H \to V/M$ is then well defined. We indicate the proof. Suppose that σ': $H \to V$ is another such section. Then:

$$\begin{aligned}(\sigma'(h), h)(\sigma(h), h)^{-1} &= (\sigma'(h), h)(-h^{-1}(\sigma(h)), h^{-1})\\ &= (\sigma'(h) - \sigma(h), 1).\end{aligned}$$

Hence the difference is in M and the map s makes perfect sense independent of the choice of σ. We also have the following result:

(5.1) PROPOSITION. *The map s satisfies the following identities*:
(i) $s(1) = 0$ (*note*: 0 *denotes the* zero coset M *in* V/M),
(ii) $s(xy) = s(x) + x \cdot s(y)$.

Proof. Firstly note that H acts on V/M because M is invariant under H. This is the action on the right-hand side of equation (ii). Equation (i) merely asserts that $(0, 1)$, the identity element, is in G. To show (ii) we need only observe that

$$(s(x), x)(s(y), y) = (s(x) + x \cdot s(y), xy).$$

We call (i) and (ii) the **cocycle identities** and s a **1-cocycle**. The set of all such 1-cocycles forms an abelian group (because V/M is) and is denoted $Z^1(H, V/M)$.

The procedure that we have just described that leads from G to the 1-cocyle s is reversible. The group G can easily be reconstructed from the 1-cocyle by writing

$$G = \{(v, h) \in \mathrm{Isom}(\mathbb{E}^n) : h \in H \quad \text{and} \quad v \in s(h)\}.$$

(Remember that $s(h)$ is a coset of M in V.) This assertion is easy to check. Hence instead of classifying space groups we are reduced to classifying 1-cocycles s: $H \to V/M$, i.e., functions

satisfying certain identities, a seemingly more manageable task. Now we can invoke the second theorem of Bieberbach. Isomorphisms of space groups are detected by conjugacy of the space groups inside the affine group $\mathrm{Aff}(\mathbb{E}^n)$. What remains to do is understand the effect of conjugating by an element of $\mathrm{Aff}(\mathbb{E}^n)$ over in the realm of 1-cocycles. This is surprisingly easy and straightforward.

Firstly each element of $\mathrm{Aff}(\mathbb{E}^n) = V \rtimes \mathrm{GL}(V)$ is the composition of a translation $(a,1)$, $a \in V$, and a linear mapping $(0, g)$, $g \in \mathrm{GL}(V)$. Hence it suffices to analyze the effect of conjugating a 1-cocycle by each of these separately.

Suppose a is an element of the vector space V and the 1-cocycle s is induced from $\sigma\colon V \to M$. Then we can compute

$$\begin{aligned}(a,1)(\sigma(h),h)(a,1)^{-1} &= (a,1)(\sigma(h),h)(-a,1)\\ &= (a+\sigma(h),h)(-a,1)\\ &= (a+\sigma(h)-h(a),h).\end{aligned}$$

Passing to the quotient $V \to V/M$, we see that conjugating by $(a,1)$ changes the 1-cocycle by adding another 1-cocycle of the form $b_a(h) = \alpha - h(\alpha)$, where α denotes a mod M. That this function really is a 1-cocycle follows from the checking that $b_a(1) = \alpha - 1(\alpha) = 0$ and

$$b_a(hh') = \alpha - hh'(\alpha) = \alpha - h(\alpha) + h(\alpha) - hh'(\alpha) = b_a(h) + h(b_a(h')).$$

The 1-cocycles of this form are called **1-coboundaries** and form a subgroup of $Z^1(H, V/M)$ denoted $B^1(H, V/M)$. From the point of view of our study of space groups these 1-coboundaries should be considered trivial, so it is natural to consider the quotient group obtained by dividing out the 1-coboundaries: $Z^1(H, V/M)/B^1(H, V/M)$. This group is usually denoted $H^1(H, V/M)$ and is called the **1-dimensional cohomology group of** H **with coefficients in** V/M.

It now remains to consider the effect of conjugating by $(0, g)$ where g is an element of $\mathrm{GL}(V)$. So we compute

$$(0,g)(\sigma(h),h)(0,g)^{-1} = (0,g)(\sigma(h),h)(0,g^{-1}) = (g(\sigma(h)), ghg^{-1}).$$

This suggests defining the following action of $\mathrm{GL}(V)$ on the group of 1-cocycles. If g is in $\mathrm{GL}(V)$ and s is in $Z^1(H, V/M)$, then

$$(g \cdot s)(h) = gs(g^{-1}hg).$$

Then the above computations can be summarized by saying that the effect of conjugating s by g in $\mathrm{GL}(V)$ is precisely gs. Hence if g is in the normalizer $N(H, M)$ of H in $\mathrm{Aut}(M)$, then the 1-cocycles should be identified. Putting this all together we obtain (see [**26**, p.35]) the main result:

(5.2) MAIN THEOREM OF MATHEMATICAL CRYSTALLOGRAPHY. *There exists a one-to-one correspondence between space groups in the arithmetic crystal class (H, M) and the orbits of $N(H, M)$ acting on the 1-dimensional cohomology group $H^1(H, V/M)$.*

REMARK (for experts). More sophisticated readers might have expected to see the 2-dimensional cohomology group $H^2(H, M)$ play the pivotal role in the classification of space groups. In fact, using the long exact sequence associated to the coefficient sequence $O \to M \to V \to V/M \to O$ we get an isomorphism: $H^2(H, M) \cong H^1(H, V/M)$.

6. An Example. Not only is Theorem (5.2) a beautiful and powerful theorem, it also gives one a computational hold on classifying crystals. We return now to the examples of the introduction and show how these "tilings" can be distinguished with the use of cohomology of groups. We begin with an easy lemma:

(6.1) LEMMA. *If $s\colon H \to V/M$ is a 1-cocycle and x is in H, then*

$$s(x^k) = (1 + x + x^2 + \cdots + x^{k-1})s(x).$$

Proof. Induct on k.

The dihedral group D_4 of order 8 is generated by elements R and S subject to the relations

$$S^4 = 1 \qquad R^2 = 1 \qquad RSR = S^{-1}.$$

We will often write this last relation as $(SR)^2 = 1$. The group D_4 admits a unique arithmetic crystal class $(D_4, \mathbb{Z}e_1 \oplus \mathbb{Z}e_2)$ where $\{e_1, e_2\}$ is the standard basis and the action $D_4 \to \mathrm{Aut}(\mathbb{Z}e_1 \oplus \mathbb{Z}e_2) = \mathrm{GL}(2, \mathbb{Z})$ is given by

$$S \mapsto \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \qquad R \mapsto \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Geometrically S is a rotation through an angle $\pi/2$ and R is a reflection through the y-axis. We are going to apply (5.2) to this arithmetic crystal class (H, M) to find the two-dimensional space groups with point group D_4.

To compute the cohomology group we begin by identifying the 1-cocycles $s\colon D_4 \to \mathbb{R}^2/\mathbb{Z}^2$. Certainly s is determined by its values on S and R. Let's suppose

$$s(S) = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \qquad s(R) = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix},$$

where we denote the elements of $\mathbb{R}^2/\mathbb{Z}^2$ by column vectors. By the first of the cocycle conditions, s sends any relation into $\mathbb{Z}^2$. Hence by (6.1)

$$\begin{aligned} s(R^2) &= (1 + R)s(R) \\ &= \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 2b_2 \end{pmatrix} \in \mathbb{Z}^2. \end{aligned}$$

So $b_2 \in (1/2)\mathbb{Z}$, a half-integer. Similarly,

$$\begin{aligned} s\big((SR)^2\big) &= (1 + SR)s(SR) \\ &= \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\right] s(SR) \\ &= \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\left[\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}\right] \\ &= \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} a_1 - b_2 \\ a_2 + b_1 \end{pmatrix} \in \mathbb{Z}^2, \end{aligned}$$

or, equivalently, $a_1 - b_2 \equiv a_2 + b_1 \pmod{\mathbb{Z}}$. Since $1 + S + S^2 + S^3 = 0$, the condition $s(S^4) \in \mathbb{Z}^2$ places no restriction.

Now we determine the coboundaries. If

$$v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \in \mathbb{R}^2,$$

then the 1-coboundary b_v is given by $b_v(g) = v - gv = (1 - g)v$. We compute b_v on the generators S, R:

$$\begin{aligned} b_v(S) &= (1 - S)v = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} v_1 + v_2 \\ v_2 - v_1 \end{pmatrix}, \\ b_v(R) &= (1 - R)v = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 2v_1 \\ 0 \end{pmatrix}. \end{aligned}$$

To compute H^1 we begin by simplifying our given 1-cocycle s by adding appropriate b_v's. If we

set

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2}(a_1 - a_2) \\ \frac{1}{2}(a_1 + a_2) \end{pmatrix},$$

then clearly $(s - b_v)(S) = 0$ and

$$(s - b_v)(R) = \begin{pmatrix} b_1 - a_1 + a_2 \\ b_2 \end{pmatrix}.$$

The cocycle condition gives us

$$b_2 \in \left(\frac{1}{2}\right)\mathbb{Z}$$

and

$$b_1 + b_2 \equiv a_1 - a_2 (\operatorname{mod} \mathbb{Z}).$$

But since $s - b_v$ is also a cocycle with $a_1 = a_2 = 0$, we get

$$b_1 + b_2 \equiv 0 (\operatorname{mod} \mathbb{Z}).$$

Hence, mod $\mathbb{Z}$, we have only two possibilities:

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \equiv \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \equiv \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}.$$

Hence $H^1(D_4, V/M) \cong \mathbb{Z}_2$, the cyclic group with two elements.

The space groups G and G' of the introduction correspond to the trivial and non-trivial elements of this $H^1 \cong \mathbb{Z}_2$. Observe that the normalizer plays no role in this calculation since the normalizer permutes the nonzero elements of the cohomology group. The reader should study Fig. 2b to convince himself that although R is not a symmetry of the "wallpaper", the element $((\frac{1}{2}, \frac{1}{2}), R) \in \operatorname{Isom}(\mathbb{E}^2)$ is.

7. Finiteness. In 1900 at the International Congress of Mathematicians in Paris, Hilbert gave a list of what he considered the outstanding, unsolved mathematical problems of the day. His inclusion of the following problem generated early interest in the mathematical approach to crystallography (see [**16**]):

> "Is there in n-dimensional Euclidean space... only a finite number of essentially different kinds of groups of motions with a [compact] fundamental domain?"

Despite Hilbert's pessimism about a rapid solution of this problem, L. Bieberbach answered it affirmatively within the following decade.

(7.1) BIEBERBACH'S THIRD THEOREM. *For each n, there are only finitely many isomorphism classes of n-dimensional space groups.*

As we have already discussed in Section 2, one can replace "isomorphism class" by "affine conjugacy class" to get an apparently stronger statement.

After Bieberbach proved his First Theorem (see 3.2 above) the solution of Hilbert's problem followed from essentially known results. The strategy of the argument follows three steps.

1. For each n, there are only finitely many n-dimensional geometric crystal classes.

2. For each geometric crystal class, there are only finitely many arithmetic crystal classes geometrically equivalently to it.
3. For each arithmetic crystal class, there are only finitely many space groups in that class.

The final step is the easiest. According to the Main Theorem of Mathematical Crystallography (5.2), it suffices to check that $H^1(H, V/M)$ is finite. This is, in fact, an elementary fact from the theory of group cohomology (see [**6**]). An elementary and direct proof can be found in Schwarzenberger's book [**26**, p. 130].

The first step was given a group-theoretic proof by Minkowski [**18**] in 1905. It also follows from the so-called Minkowski-Siegel reduction theory [**17**] for positive-definite quadratic forms. A readable proof can be found in [**19**].

Finally the second step in the proof is a special case of the Jordan-Zassenhaus theorem of representation theory, a theorem we have already discussed in Section 4.

In 1948, Zassenhaus [**32**] observed that this proof could be turned into an effective algorithm for enumerating space groups. It was only in 1976 that this algorithm was fully implemented in dimension 4 and generated a complete list (with much additional data) of the 4783 (!) four-dimensional space groups.

The first step of the above strategy requires a listing of the conjugacy classes of finite subgroups of $\mathrm{GL}(n, \mathbb{Z})$. One begins this enumeration by finding the maximal finite subgroups of $\mathrm{GL}(n, \mathbb{Z})$. In fact, this has been worked out for $n \leqslant 7$ in the work of Dade [**9**] and Plesken-Pohst [**22**], [**23**]. All of the conjugacy classes of finite subgroups (i.e., arithmetic crystal classes) can then be found by applying certain "subgroup subroutines" to the list of maximal ones; this provides a count of the number of geometric crystal classes. Steps two and three then require writing down the integral representations of the groups, computing cohomology and normalizers and finally the set of orbits. More details on these procedures can be found in [**5**] as well as extensive computer printouts of the results.

Here is a table of some of the known statistics:

TABLE 1

dimension	# geometric crystal classes	# arithmetic crystal classes	# space groups
2	10	13	17
3	32	73	219
4	227	710	4783

Finally we mention that Schwarzenberger [**27**] (see also [**26**, p. 96]) has shown that if s_n denotes the number of space groups in dimension n, then s_n grows at least as fast as 2^{n^2} and conjectures that this is the exact asymptotic result.

8. Wallpaper. We began with our intuition about crystals in three-dimensional space. Historically this was also the starting point of mathematical crystallography. The possible point groups for the three-dimensional space groups were first determined by Hessel (1829); a modern readable account can be found in [**2**, Theorem 2.5.2.]. In an apparently surprising coincidence (but see [**26**, p. 132]) the crystals in three-dimensional space were classified independently and almost simultaneously by Fedorov (in Russia), Schoenflies (in Germany) and Barlow (in England) in the latter part of the nineteenth century. Their work built upon earlier contributions of Hessel, Bravais, Möbius, Jordan, and Sohncke. The methods they employed were *ad hoc* and directly geometric. A modern cohomological approach to the classification of the 219 crystals in the spirit of the techniques we have discussed can be found in [**26**] and [**14**].

REMARK. It should be mentioned that most crytallographers actually insist that there are 230 crystals. This discrepancy arises from 11 so-called enantiomorphic pairs—as in organic chemistry. These crystals differ by a mirror reflection (i.e., one is "left-handed" and the other

"right-handed"). The mathematical explanation is that crystallographers use for their notion of equivalence of space groups the stronger one of conjugacy inside the special affine group $S\mathrm{Aff}(\mathbb{E}^n) = V \rtimes \mathrm{GL}^+(n,\mathbb{R})$, i.e., affine mappings with positive determinant (hence omitting the mirror reflections with determinant -1). A completely mathematical description of this phenomenon has been worked out by Maxwell [**15**].

Somewhat later it was realized that the same methods could be applied in the easier cases of dimensions 1 and 2. In dimension one, there are only three space groups contained in $\mathrm{Isom}(\mathbb{E}^1) = \mathbb{R} \rtimes \mathbb{Z}_2$. (What are they?) There are also 7 "frieze" patterns that can be viewed as the linear patterns that wind around Grecian urns. (Try finding 7 representative patterns and convincing yourself that these are all of them; see [**13**].)

The space groups in dimension 2 are usually called the **wallpaper groups**. There are 17 of them and they were first catalogued by Pólya [**24**] and Niggli [**20**] in the twenties. A completely elementary exposition of these results can be found in [**28**] (see also [**26**, Chap. 1]).

The reader is encouraged to apply the techniques of Sections 4–5 to check cohomologically the table of wallpaper groups given below. It is also instructive to compare this table with those that appear in Schattschneider's article [**25**].

The fact that the groups listed in the first column below are the only possible point groups is a fact usually attributed to Leonardo Da Vinci (for example, see Weyl [**29**]). For an elementary proof see [**2**].

TABLE 2. Wallpaper groups.

point group	acc	# $H^1(H, V/M)$	# space groups	notation
e	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$	1	1	p1
$\mathbb{Z}_2$	$\begin{pmatrix}-1&0\\0&-1\end{pmatrix}$	1	1	p2
$\mathbb{Z}_3$	$s_3=\begin{pmatrix}0&-1\\1&-1\end{pmatrix}$	1	1	p3
$\mathbb{Z}_4$	$s_4=\begin{pmatrix}0&-1\\1&0\end{pmatrix}$	1	1	p4
$\mathbb{Z}_6$	$s_6=\begin{pmatrix}0&1\\1&1\end{pmatrix}$	1	1	p6
D_1	$r=\begin{pmatrix}-1&0\\0&1\end{pmatrix}$	2	2	pm, pg
	$\begin{pmatrix}1&1\\0&-1\end{pmatrix}$	1	1	cm
D_2	$\pm\begin{pmatrix}-1&0\\0&1\end{pmatrix}$	4	3	pmm, pmg, pgg
	$\pm\begin{pmatrix}1&1\\0&-1\end{pmatrix}$	1	1	c2mm
D_3	$r \quad s_3$	1	1	p3m1
	$\begin{pmatrix}-1&0\\1&1\end{pmatrix}\begin{pmatrix}-1&-1\\1&0\end{pmatrix}$	1	1	p31m
D_4	$r \quad s_4$	2	2	p4m, p4g
D_6	$r \quad s_6$	1	1	p6m

The book of Schwarzenberger [**26**] is a natural source for further details and elaborations of the results described here. The more ambitious reader might consult the book of Wolf [**30**] where the ideas of crystallography are extended to the other non-Euclidean geometries—spherical and hyperbolic. The spherical case is fairly well understood (although the required group-theoretic labors were substantial) while the hyperbolic case is still something of a mystery. The two-dimensional hyperbolic situation was studied extensively by Fricke and Klein [**10**] in the latter part of the nineteenth century in the context of their work on automorphic functions. The reader should browse through the book of Magnus [**12**] (or a catalogue of Escher drawings) to see what wallpaper in hyperbolic 2-space looks like.

9. Epilogue: Flat Manifolds. In Section 3 we introduced the fundamental notion of a space group by examining the quotient space $\mathbb{E}^n/G$. If we consider the simple case of a cyclic point group, say $\mathbb{Z}^2 \rtimes \mathbb{Z}_3$ acting on $\mathbb{E}^2$ by a 120° rotation, the resulting quotient space $\mathbb{E}^2/\mathbb{Z}^2 \rtimes \mathbb{Z}_3$ has a nasty singularity (or corner) at the origin.

Mathematicians generally find such singularities unpleasant so it is natural to ask whether there are space groups G for which the quotient $\mathbb{E}^n/G$ is without singularities, i.e., is a **manifold**. We call such a group a **Bieberbach group**. We already saw such an example back in Section 3 where we considered $\mathbb{Z} \oplus \mathbb{Z}$ acting by translations on $\mathbb{E}^2$. The quotient was seen to be a **torus** which is a compact surface, a two-dimensional manifold. Of course this example generalizes to any dimension, since $\mathbb{E}^n/\mathbb{Z}^n$ is an n-dimensional torus. In this family of examples the point group H (now called the **holonomy group**) is trivial. In general one has that $\mathbb{E}^n/G$ is a manifold if and only if no element of G fixes a point in $\mathbb{E}^n$. In such a case we call the action of G on $\mathbb{E}^n$ a **free** action. For example, a pure rotation could not be in G if we want $\mathbb{E}^n/G$ to be a manifold. In fact one has the following general algebraic criterion:

(9.1) PROPOSITION. *A space group $G \subset \mathrm{Isom}(\mathbb{E}^n)$ is a* **Bieberbach group** *if and only if it is torsion-free.*

Recall that a group is **torsion-free** if it has no elements of finite order. In particular, a Bieberbach group is never a semi-direct product, because in that case the holonomy group H would inject into G via any map that splits G (see Section 2). In fact we can re-express the criterion of (9.1) by saying that the short exact sequence defining G does not split over any subgroup of H. This can then be expressed as a condition on the 1-dimensional cohomology class defining G.

The manifolds that arise as quotients of Euclidean space have certain special differential-geometric properties. They are precisely the **flat manifolds**. These are Riemannian manifolds that have zero curvature. (For the precise definition of curvature see [**21**]. They are also the Riemannian manifolds whose Riemannian universal covering spaces are Euclidean space $\mathbb{E}^n$.) In dimension two there are precisely 2 flat manifolds: the torus and the Klein bottle. The second example arises from the unique non-trivial class in $H^1(\mathbb{Z}_2, \mathbb{R}^2/\mathbb{Z}^2)$, where $\mathbb{Z}_2$ acts by flipping the factors (it is not orientable). In dimension 3 there are 10 flat manifolds among the 219 crystals, and in dimension 4 there are 75 flat manifolds among the 4783 crystals. It is unknown how many 5-dimensional flat manifolds there are, but Schwarzenberger [**26**] has shown that there are at least 9806 space groups with point group $(\mathbb{Z}_2)^5$.

References

1. A. Beardon, The Geometry of Discrete Groups, Springer-Verlag, New York, 1983.

2. C. Benson and R. Grove, Finite Reflection Groups, Bogden and Quigley, Tarrytown-on-Hudson, New York, 1971.

3. L. Bieberbach, Über die Bewegungsgruppen der Euklidischen Räume I, Math. Ann., 70 (1910) 297–336.

4. ______, Über die Bewegungsgruppen der Euklidischen Räume II, Math. Ann., 72 (1912) 400–412.

5. H. Brown, R. Bülow, J. Neubüser, H. Wondratschek, and H. Zassenhaus, Crystallographic Groups of Four-Dimensional Space, Wiley, New York, 1978.

6. K. Brown, Cohomology of Groups, Springer-Verlag, New York, 1982.

7. H. Buser and H. Karcher, The Bieberbach case in Gromov's Almost Flat Manifold Theorem, in Global Differential Geometry and Global Analysis (Berlin 1979), Lecture Notes in Mathematics # 838, Springer-Verlag, New York, 1981.

8. C. Curtis and I. Reiner, Representation Theory of Finite Groups and Associative Algebras, Interscience, New York, 1962.

9. E. C. Dade, The maximal finite groups of 4 × 4 matrices, Illinois J. Math., 9 (1965) 99–122.

10. R. Fricke and F. Klein, Vorlesungen über die Theorie der automorphen Functionen, vol. 1, Leipzig, 1887.

11. G. Frobenius, Über die unzerlegbaen diskreten Bewegungsgruppen, Sitzungsber. Preuss. Akad. Wiss. Phys. Math. Kl., (1911) 654–655.

12. W. Magnus, Non-euclidean Tesselations and their Groups, Academic Press, New York, 1976.

13. G. Martin, Transformation Geometry: An Introduction to Symmetry, Springer-Verlag, New York, 1982.

14. G. Maxwell, The crystallography of Coxeter groups, J. Algebra, 25 (1975) 159–177.

15. ______, On the automorphism groups of space groups, Comm. Algebra, 6(14) (1978) 1489–1496.

16. J. Milnor, Hilbert's Problem 18: On crystallographic groups, fundamental domains and on sphere-packing, in Proc. Symp. Pure Math. XXVIII, Amer. Math. Soc., Providence, RI, 1976.

17. H. Minkowski, Zur Theorie der positiven quadratischen Formen, Crelle J. reine u. angew. Math., 106 (1890) 5-26.

18. ______, Diskontinuitätsbereiche fur arithmetische Äquivalenz, J. Reine Angew. Math., 129 (1905) 220-274.

19. G. Mostow, Rigidity of Locally Symmetric Spaces, Annals of Math. Studies no. 78, Princeton University Press, Princeton, NJ, 1973.

20. P. Niggli, Die Flächensymmetrien homogener Diskontinuen, Z. Krist., 60 (1924) 183–298.

21. B. O'Neill, Elementary Differential Geometry, Academic Press, New York, 1966.

22. W. Plesken and M. Pohst. On maximal finite irreducible subgroups of GL(n, $\mathbb{Z}$) I, The five and seven dimensional cases, Math. Comput. 31 (1977) 536–551.

23. ______, On maximal finite irreducible subgroups of GL(n, $\mathbb{Z}$) II, The six dimensional case, Math. Comput., 31 (1977) 552–573.

24. G. Pólya, Über die Analogie der Kristallsymmetrie in der Ebene, Z. Krist., 60 (1924) 278–282.

25. D. Schattschneider, The plane symmetry groups: their recognition and their notation, this MONTHLY, 85 (1978) 439–450.

26. R. L. E. Schwarzenberger, N-dimensional Crystallography, Research Notes in Mathematics no. 41, Pitman Publishing, London, 1980.

27. ______, Crystallography in spaces of arbitrary dimension, Proc. Cambridge Philos. Soc., 76 (1974) 23–32.

28. ______, The 17 plane symmetry groups, Math. Gaz., 58 (1974) 123–131.

29. H. Weyl, Symmetry, Princeton University Press, Princeton, NJ, 1952.

30. J. Wolf, Spaces of Constant Curvature, McGraw-Hill, New York, 1967.

31. H. Zassenhaus, Beweis eines Satzes über diskrete Gruppen, Abh. Math. Sem. Univ. Hamburg, 12 (1938), 2890312.

32. ______, Über einen Algorithmus zur Bestimmung der Raumgruppen, Comment. Math. Helv., 21 (1948) 117–141.

THE PARTIAL ORDER OF ITERATED EXPONENTIALS

BARRY W. BRUNSON

Department of Mathematics, Western Kentucky University, Bowling Green, KY 42101

1. Introduction. Which is larger, π^e or e^π? Before the advent of hand calculators, this was a more interesting exercise in calculus texts than at present. But even with current technology, try finding the larger of

$$e^{e^{\pi^{e^{\pi^{\pi}}}}} \quad \text{and} \quad e^{\pi^{\pi^{e^{\pi^{e}}}}}$$

by computing the decimal approximations.

Here is a more general question: let $e \leq x_1 \leq x_2 \cdots \leq x_n$, and for each permutation $(y, z, \dots, w)$ of the x_i's, let $T(y, z, \dots, w)$ be the "iterated exponential":

$$T(y, z, \dots, w) = y^{z^{\cdot^{\cdot^{\cdot^{w}}}}}.$$

Such an expression is called a *tower* with levels $1, 2, \dots, n$; level 1 is at the base, and association is always to the upper right:

Barry Brunson received the Ph.D. from Indiana University under the direction of Victor Goodman. His mathematical interests are broad, and include martingales with unusual index sets, analysis, foundations, and combinatorics. Among perceived ills of the world, he sees the misuse and abuse of statistics as causing much social harm. He entered both the professoriate and parenthood 18 years after high school, and spends considerable time with children Cory and Nicole, whose ages have a sum of five and a product of four.

$$T(y, z, w) = y^{(z^w)}.$$

Given two such permutations $\mathbf{y} = (y_1, \dots, y_n)$ and $\mathbf{z} = (z_1, \dots, z_n)$, which is larger, $T(\mathbf{y})$ or $T(\mathbf{z})$?

The basic result of this paper, Theorem 1, shows that the answer often depends only on the permutations and not on the specific values of the x_i's. This leads to a partial order on the set of towers (identified with the permutations) and, by implication, on the symmetric group S_n itself. A complete characterization of when one element covers another remains elusive, but appears to have an interesting connection with a certain set of generators of S_n.

Informally, Theorem 1 states that, given two permutations which differ only by the transposition of two elements x_i and x_{i+j}, the larger tower is the one having x_{i+j} "farther from the ground." This generalizes to exponentiations a remark of Hardy, Littlewood, and Pólya, made in the context of sums of products of numbers [**HLP**, p. 262]:

> "To get the maximum statical moment with respect to an end of the rod, we hang the heaviest weights on the hooks farthest from that end."

Among the immediate consequences are that the largest tower has its components arranged in ascending order, and the smallest in descending order. Of course, for specific values of the variables, the $n!$ possible towers are just numbers and hence totally ordered. Determining that order may be no mean task, however. The smallest tower with $n = 4$ levels exceeds 10^{10^6}; the smallest with $n = 5$ dwarfs Kasner and Newman's googolplex$(= 10^{10^{100}})$, and that with $n = 6$ makes Skewes' number $(= 10^{10^{10^{34}}})$ miniscule by comparison. Direct computation is out of the question, although estimation with logarithms can be of some use.

This inquiry was stimulated by Problem 271 in the College Mathematics Journal (March '84), posed by Robert E. Shafer, which asked for the largest member of the set

$$\left\{ x^{y^z}, x^{z^y}, y^{x^z}, y^{z^x}, z^{x^y}, z^{y^x} \right\}, \quad \text{where } e \le x \le y \le z.$$

2. The basic result. For $n \ge 1$, let

$$R_e^n = \left\{ \mathbf{x} = (x_1, x_2, \dots, x_n) \colon x_i \ge e \right\}.$$

We consider only towers $T(\mathbf{x})$ defined for vectors $\mathbf{x}$ in R_e^n.

Since towers are well defined (and real) for arbitrary positive components, one may wonder about the restriction that the components be at least e. One practical reason is because things get very complicated otherwise. The simple case of x^y versus y^x, for $x \le y$, illustrates the difficulties. If $0 < x \le 1$, then $x^y \le y^x$ for all y, and $x^y \ge y^x$ if $x \ge e$. But if $1 < x < e$, then $x^y < y^x$ for $y < y_0$ and the inequality is reversed for $y > y_0$, where y_0 depends on x (of the two solutions to the equation $\ln y/y = \ln x/x$ in y, y_0 is the one which exceeds e).

The statement of the main result is more verbose than the informal description above, but is notationally more amenable to proof.

THEOREM 1. *Let* $(z_1, z_2, \dots, z_n, x, y) \in R_e^{n+2}$, $n \ge 0$, *with* $x \le y$. *Then for any appropriate* i *and* j,

$$T(z_1, \dots, z_i, x, z_{i+1}, \dots, z_j, y, z_{j+1}, \dots, z_n) \ge T(z_1, \dots, z_i, y, z_{i+1}, \dots, z_j, x, z_{j+1}, \dots, z_n),$$

and equality holds if and only if $x = y$.

We may consider 0 and n "appropriate" values for i and/or j, as well as $i = j$, by natural interpretation; for instance, $i = 0$, $j = n$ indicates the inequality

$$T(x, z_1, \dots, z_n, y) \ge T(y, z_1, \dots, z_n, x).$$

The proof is elementary but requires attention to several cases. We begin by establishing a lemma of independent interest, the proof of which gives the flavor of the general technique. Recall that to show that $a \ge b$, it suffices to show either that $\ln a - \ln b \ge 0$, or that $\ln\ln a - \ln\ln b \ge 0$, provided each term is defined; we will routinely use this device, together with the

obvious fact that $T(\mathbf{x})$ is monotonically increasing in each argument.

LEMMA 2. *Let $e \le x \le y$ and $C \ge 1$. Then $x^{y^C} \ge y^{x^C}$, with equality if and only if $x = y$.*

Proof. Fix x, and let

$$g(y) = \ln\ln(x^{y^C}) - \ln\ln(y^{x^C}) = C\ln y + \ln\ln x - C\ln x - \ln\ln y, \quad \text{for } y \ge x.$$

Then $g(x) = 0$, and $g'(y) = C/y - 1/(y\ln y)$, which is nonnegative (and positive unless $x = y = e$, $C = 1$). □

LEMMA 3. *Let $f(y) = T(x_1, x_2, \dots, x_n, y)$. Then*

$$f'(y) = \left[\prod_{k=1}^{n} T(x_k, x_{k+1}, \dots, x_n, y)\right] \prod_{i=1}^{n} \ln x_i.$$

Proof. Use logarithmic differentiation and induction on n. □

Proof of Theorem 1. Observe that if we can prove the theorem for each case in which $i = 0$, then the corresponding case in which $i > 0$ will follow from the monotonicity of T; the desired inequality will be of the form

$$T(z_1, \dots, z_i, T_{xy}) \ge T(z_1, \dots, z_i, T_{yx}),$$

where $T_{xy} \ge T_{yx}$ is an inequality already established. So assume that $i = 0$. We have three slightly different cases: $j = 0$, $j = n$, and $0 < j < n$.

If $j = 0$, the desired inequality is

$$T(x, y, z_1, \dots, z_n) \ge T(y, x, z_1, \dots, z_n),$$

which follows from Lemma 2, with $C = T(z_1, \dots, z_n)$.

If $j = n$, the indicated inequality is

$$T(x, z_1, \dots, z_n, y) \ge T(y, z_1, \dots, z_n, x).$$

Fix x, and let $f(y)$ be the difference of the logarithms:

$$f(y) = T(z_1, \dots, z_n, y)\ln x - T(z_1, \dots, z_n, x)\ln y, \quad \text{for } y \ge x.$$

Then $f(x) = 0$, and by Lemma 3,

$$f'(y) = \left[\prod_{k=1}^{n} T(z_k, z_{k+1}, \dots, z_n, y)\right]\left(\prod_{i=1}^{n} \ln z_i\right)\ln x - T(z_1, \dots, z_n, x)/y.$$

Now each factor in the first term above is at least unity, and replacing y by x in any tower could not increase its size. Thus,

$$f'(y) \ge T(z_1, \dots, z_n, x)[1 - 1/y],$$

which is positive.

Note that all possibilities when $n = 1$ have been dealt with, so assume that $n \ge 2$, and that the theorem holds for fewer than n z's.

Suppose now that $0 < j < n$. We need to show that

$$T(x, z_1, \dots, z_j, y, z_{j+1}, \dots, z_n) \ge T(y, z_1, \dots, z_j, x, z_{j+1}, \dots, z_n).$$

Fix x, and let $f(y)$ be the difference of the iterated logarithms, for $y \ge x$. Since the cases $j = 1$ and $j > 1$ require but marginally different treatments ($f'(y)$ is easier to compute when $j = 1$), we present only the proof for $j > 1$. Note that granting the omitted proof permits us to henceforth assume that $n \ge 3$ (and that the theorem holds for fewer than n z's). We have

$$\begin{aligned} f(y) = {} & T(z_2, \dots, z_j, y, z_{j+1}, \dots, z_n)\ln z_1 + \ln\ln x \\ & - T(z_2, \dots, z_j, x, z_{j+1}, \dots, z_n)\ln z_1 - \ln\ln y, \end{aligned}$$

for $y \geq x$. Then $f(x) = 0$, and by Lemma 3 and the chain rule,

$$f'(y) = \prod_{k=2}^{j} T(z_k, z_{k+1}, \ldots, z_j, y, z_{j+1}, \ldots, z_n)$$
$$\times T(z_{j+1}, \ldots, z_n)\, y^{T(z_{j+1}, \ldots, z_n)-1}$$
$$\times \prod_{i=1}^{j} \ln z_i - \frac{1}{y \ln y}.$$

Now, $y^{T(z_{j+1}, \ldots, z_n)-1} \geq y^{e-1}$, and every other factor in the first term is at least unity. Thus,

$$f'(y) \geq \frac{1}{y}\left[y^e - \frac{1}{\ln y}\right],$$

which is positive. □

REMARK. The case $j = n$ above is a consequence of Lemmas 1 and 2 of the published solution (November '85) of College Mathematics Journal Problem 271, due to Allen Kaufman. Furthermore, an alternative proof of the case $0 < j < n$ could be similarly derived by extending that Lemma 1 to prove that $T(z_1, \ldots, z_j, y, z_{j+1}, \ldots, z_n)/y$ is an increasing function of y.

We are provided a method for determining the relative size of *some* towers. If one tower can be obtained from another by a sequence of transpositions, such that each one results in a larger number moving closer to the base, then the original tower is larger; e.g., for the example at the beginning,

$$e^{e^{\pi^{e^{\pi^{\pi}}}}} > e^{\pi^{e^{e^{\pi^{\pi}}}}} > e^{\pi^{e^{\pi^{e^{\pi}}}}}.$$

3. The partial order. Theorem 1 leads directly to a partial order $(S_n, \prec)$ on the symmetric group in the following sense. If $\mathbf{x} = (x_1, \ldots, x_n)$, and $\sigma \in S_n$, let $\sigma(\mathbf{x}) = (x_{\sigma(1)}, \ldots, x_{\sigma(n)})$.

DEFINITION. (1) A vector $\mathbf{x} = (x_1, \ldots, x_n)$ in R_e^n is called *increasing* if $x_1 \leq x_2 \leq \cdots \leq x_n$. (2) If $\sigma, \tau \in S_n$, then $\sigma \prec \tau$ if $T(\sigma(\mathbf{x})) \leq T(\tau(\mathbf{x}))$ for every increasing $\mathbf{x}$ in R_e^n.

Henceforth the word "tower" will play a triple role: as one of the $n!$ functions $T(\sigma(\mathbf{x}))$, for $\sigma \in S_n$; as the evaluation of such for a specific vector $\mathbf{x}$; and occasionally, as an object to be identified with the permutation σ. Given context, no ambiguity should result.

A partial order is a binary relation which is reflexive, antisymmetric, and transitive. The only nontrivial property of the relation $\prec$ is antisymmetry; that is, does the fact that $T(\sigma(\mathbf{x})) = T(\tau(\mathbf{x}))$ for all increasing $\mathbf{x}$ imply that $\sigma = \tau$? This does seem intuitively obvious, but we acknowledge the fragility of intuition by sketching a proof.

PROPOSITION 4. *Suppose that* $T(\sigma(\mathbf{x})) = T(\tau(\mathbf{x}))$ *for every increasing* $\mathbf{x}$ *in* R_e^n. *Then* $\sigma = \tau$.

Proof. We prove the contrapositive; let $\sigma \neq \tau$. Suppose first that $\sigma^{-1}(n) \neq \tau^{-1}(n)$; we assume without loss of generality that $\sigma^{-1}(n) < \tau^{-1}(n)$. Consider the vector $\mathbf{x}$ consisting of $n-1$ e's followed by α, where $\alpha > e$. Then each of $T(\sigma(\mathbf{x}))$ and $T(\tau(\mathbf{x}))$ is a tower containing one α and the rest e's, and the level of α in $T(\sigma(\mathbf{x}))$ is smaller than in $T(\tau(\mathbf{x}))$. Evidently, the two towers differ only by a transposition of α and e, and by Theorem 1, $T(\sigma(\mathbf{x})) < T(\tau(\mathbf{x}))$.

In the general case, let k, $2 \leq k \leq n$, be the largest integer for which $\sigma^{-1}(k) \neq \tau^{-1}(k)$. Consider the vector $\mathbf{x}$ consisting of $k-1$ e's followed by $n-k+1$ α's, where $\alpha > e$. Again, $T(\sigma(\mathbf{x}))$ and $T(\tau(\mathbf{x}))$ differ only by a transposition of α and e, and by Theorem 1, the tower having α at the lower level is strictly smaller. Note that two distinct permutations in S_n must differ in at least two places. □

Figs. 1 and 2 show the Hasse diagrams for $n = 3$, in terms of the actual exponentials when $e \leq x \leq y \leq z$, and in terms of S_3, respectively. In Fig. 2, numerals in parentheses describe the

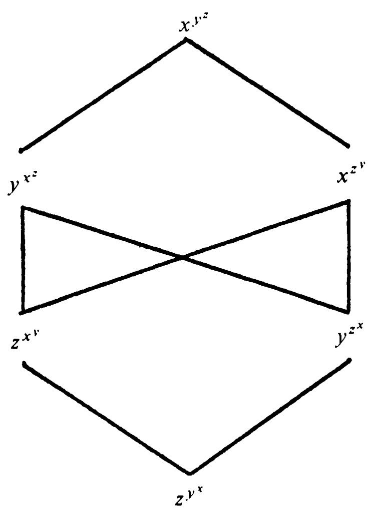


FIG. 1. Partially ordered iterated exponentials.

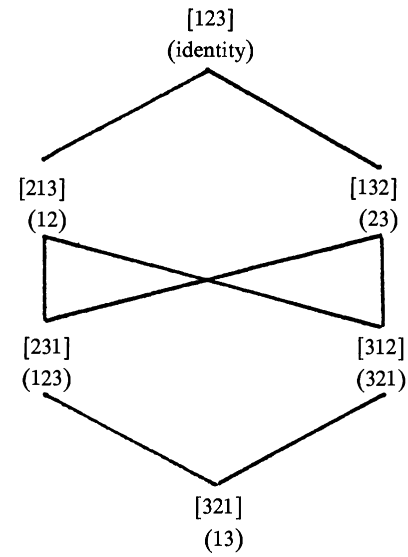


FIG. 2. $(S_3, \prec)$.

cycle decomposition, e.g., (321) represents $\tau \in S_3$ for which $\tau(3) = 2$, $\tau(2) = 1$, and $\tau(1) = 3$. As for the square brackets, the same permutation τ is given as [312], which represents $[\tau(1)\ \tau(2)\ \tau(3)]$; it constitutes the "essential" part of the usual two-line representation $\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$. The availability of both notations will be convenient in the sequel.

Here are examples which illustrate the incomparability of y^{x^z} and x^{z^y}. Let $x = y = e$, $z = e^2$; then $y^{x^z} > x^{z^y}$ since $e^2 > 2e$. Let $x = e$, $y = z = e^2$; then $y^{x^z} < x^{z^y}$ since $e^2 + \ln 2 < 2e^2$. These same values show that z^{x^y} and y^{z^x} are incomparable.

4. Coverings. When does one element of S_n *cover* another (i.e., nothing is in between)? A characterization in terms of the group operations in S_n would be nice. Transpositions would seem to result in "minimal" change, but the fact that (13) is the least element in S_3, and the identity is, as always, the greatest, indicates a more subtle relationship.

It is well known that the transpositions of the form $(i, i+1)$, $1 \le i \le n-1$, generate S_n. We call this set C_n, and show that it plays a key role.

THEOREM 5. *Let $\sigma, \tau \in S_n$, and let $\rho \in C_n$ be such that $\sigma = \tau\rho$. Then one of σ and τ covers the other.*

Proof. The statement is trivial if $n = 2$, so we assume that $n \ge 3$. Theorem 1 guarantees that σ and τ are comparable; and $\sigma = \tau\rho$ is equivalent to $\tau = \sigma\rho$ since a transposition is its own inverse. So we assume without loss of generality that $\sigma \prec \tau$.

Let $\rho = (i, i+1)$. For any increasing vector $\mathbf{x} \in R_e^n$, $T(\sigma(\mathbf{x}))$ and $T(\tau(\mathbf{x}))$ differ only in the levels of x_i and x_{i+1}. Suppose that $\xi \in S_n$ is such that $\sigma \prec \xi \prec \tau$, and let $\alpha > e$. If $i > 1$, consider $\mathbf{x} = (e, \alpha, \dots, \alpha)$. Since $x_i = x_{i+1}$, $T(\sigma(\mathbf{x})) = T(\tau(\mathbf{x})) = T(\xi(\mathbf{x}))$; but $T(\xi(\mathbf{x}))$ can differ from the others by at most a transposition of e and α. By the strict inequality aspect of Theorem 1, this requires that $\sigma^{-1}(1) = \tau^{-1}(1) = \xi^{-1}(1)$.

Proceeding inductively, assume that

$$\sigma^{-1}(j) = \tau^{-1}(j) = \xi^{-1}(j) \quad \text{for } 1 \le j \le j_0.$$

If $i > j_0 + 1$, then consider the vector $\mathbf{x}$ consisting of $j_0 + 1$ e's followed by $n - j_0 - 1$ α's. Again $x_i = x_{i+1}$, so that

$$T(\sigma(\mathbf{x})) = T(\tau(\mathbf{x})) = T(\xi(\mathbf{x})).$$

Since the levels of the first j_0 e's in $T(\xi(\mathbf{x}))$ are identical to those in the other towers, the level of the last e must be also. Thus

$$\sigma^{-1}(j) = \tau^{-1}(j) = \xi^{-1}(j) \quad \text{for } 1 \leq j < i.$$

A similar argument applies for $i + 1 < j \leq n$, and handles the case when $i = 1$. But then $\xi^{-1}(i)$ must be either $\sigma^{-1}(i)$ or $\sigma^{-1}(i + 1)$. In the former case, $\xi^{-1}(i + 1) = \sigma^{-1}(i + 1)$ as well, so that $\xi = \sigma$; in the latter case, $\xi = \tau$. □

THEOREM 6. *Let $\sigma, \tau \in S_n$ and let $\rho \in C_n$ be such that $\sigma = \rho\tau$. Then one of σ and τ covers the other.*

Proof. The technique parallels that of Theorem 5, and the details are omitted. □

5. Irreducible representations. We have just seen, in Theorems 5 and 6, that multiplying on the right or on the left by a transposition in C_n results in an immediate neighbor. This corresponds to transposing adjacent vector components or transposing components at adjacent levels in the tower, respectively. But in Fig. 4, we notice for instance that the permutation [2143] covers both [4123] and [2341], which fall in neither category. Looking at the permutations expressed as products of elements of C_n, however, reveals the following correspondence:

$$\begin{aligned}
[2143] &\to (12)(34) = (34)(12)\\
[4123] &\to (12)(23)(34)\\
[2341] &\to (34)(23)(12)
\end{aligned}$$

This, together with similar evidence, results in some interesting conjectures. But first some more terminology.

Any permutation τ in S_n can be expressed as a product of members of C_n in infinitely many ways, of course. But there is a minimal number $k = d(\tau)$, such that τ can be written as the product of k members of C_n, but not of fewer than k. Suppose that $\tau = a_1 a_2 \cdots a_k$, where k is

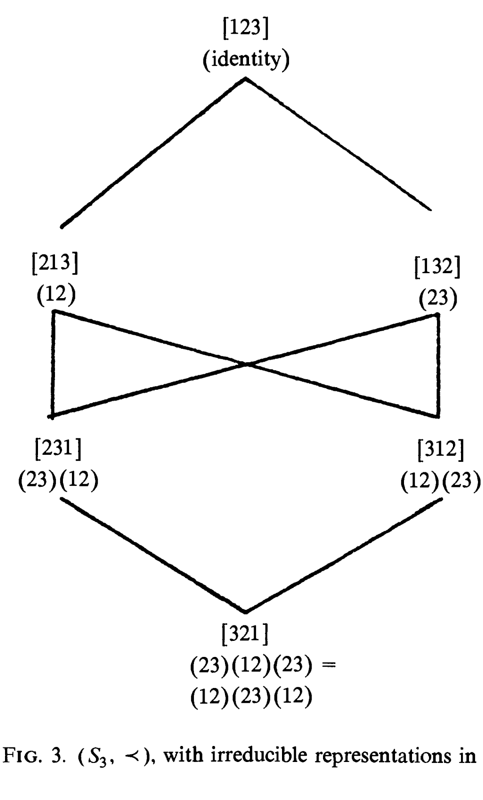


FIG. 3. $(S_3, \prec)$, with irreducible representations in C_3.

as just described. We will call $a_1 a_2 \cdots a_k$ an *irreducible* representation of τ in C_n. Irreducible representations may or may not be unique. For instance, the identity permutation has $k = 0$ and is unique, as is $\tau = a$ for $a \in C_n$. But $\tau = (13)$ can be written as either (12) (23) (12) or as (23) (12) (23).

CONJECTURE A. *Let $\tau \in S_n$, let $a_1 a_2 \cdots a_k$ be an irreducible representation of τ in C_n, and let $\sigma = a_1 a_2 \cdots a_j b a_{j+1} \cdots a_k$, where $b \in C_n$. Then the given representation of σ is irreducible, or reducible only to a representation with $k - 1$ factors, if and only if one of σ, τ covers the other.*

Figs. 3 and 4 give the Hasse diagrams for $(S_3, \prec)$ and $(S_4, \prec)$, respectively, together with

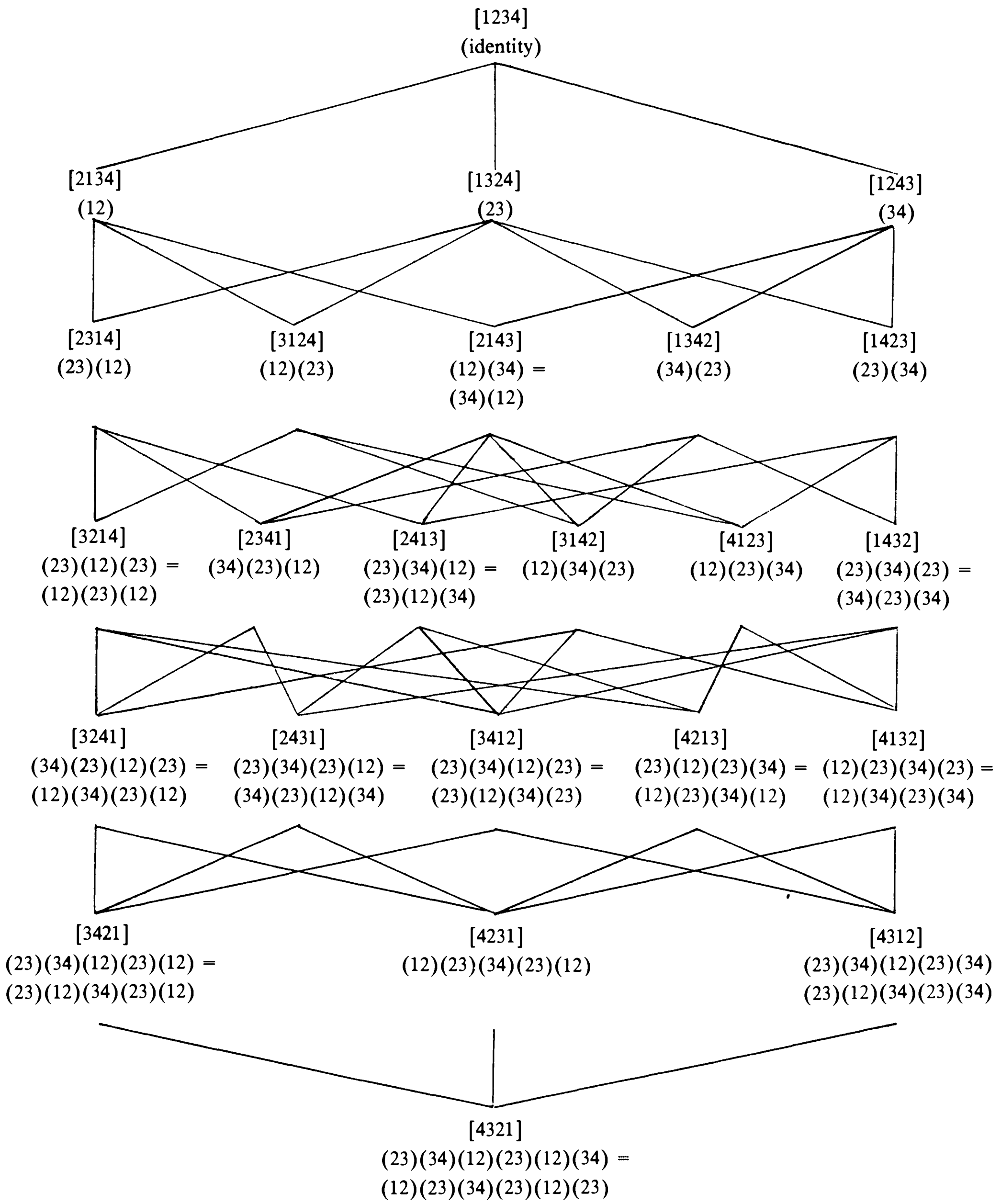


FIG. 4. $(S_4, \prec)$, with irreducible representations in C_4.

irreducible representations in C_n. An intriguing pattern is apparent. It leads to another conjecture which would, among other things, resolve the question of which of σ, τ covers the other in Theorems 5 and 6 and Conjecture A.

CONJECTURE B. *Let $d(\tau)$ be the number of factors in an irreducible representation of τ in C_n. Then $\sigma \prec \tau$ only if $d(\sigma) \geq d(\tau)$, and τ covers σ only if $d(\sigma) = d(\tau) + 1$.*

The notation $d(\tau)$ was chosen to represent the "depth" of τ. An immediate consequence of Conjecture B is that $(S_n, \prec)$ is graded by its height function, and hence satisfies the Jordan-Dedekind chain condition (cf. [**B**], p. 5).

Acknowledgment. The author wishes to thank Kyle Wallace for several stimulating conversations related to this work.

References

[B]. G. Birkhoff, Lattice Theory, 3rd ed., Providence, 1979.
[HLP]. G. H. Hardy, J. E. Littlewood, and G. Pólya, Inequalities, 2nd ed., Cambridge, 1959.

THE LAPLACE TRANSFORM INVERSION BY INSPECTION

ABRAHAM UNGAR
Department of Mathematics, North Dakota State University, Fargo, ND 58105

The theory of Laplace transforms, also referred to as operational calculus, is useful in solving some boundary value problems where the solution is determined by inversion of the operational solution which can be obtained from the operational problem. The inversion of the operational solution may be a difficult task. Surprisingly, however, there are some involved boundary value problems which are of interest and for which the Laplace transform inversion can be done by inspection.

The power of the Laplace transform inversion by inspection is demonstrated in this article by employing it for the transformation of the **separable** solution

$$\phi = \gamma e^{-s(\alpha x + i\beta y)}, \quad \beta^2 - \alpha^2 = 1, \quad \alpha, \beta, \gamma = \text{constant}, \tag{1}$$

of the reduced Laplace equation

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + s^2 \phi = 0, \tag{2}$$

into a **nonseparable** solution,

$$\psi = \frac{1}{\sqrt{r^2 + t^2}} F\left(\sinh^{-1}\left(\frac{t}{r}\right) + i\theta\right), \quad r^2 = x^2 + y^2, \quad \theta = \tan^{-1}(y/x),$$

of the Laplace equation

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial t^2} = 0, \tag{3}$$

Abraham Ungar: My special predilection for the study of wave propagation by means of integral transforms and PDEs dates from my Ph.D. in mathematical geophysics (1973) from Tel-Aviv University. My B.Sc. (1965) and M.Sc. (1967) in mathematics and physics are from The Hebrew University of Jerusalem. I was a faculty member at Rhodes University, South Africa (1978–1983), and visiting associate professor at Simon Fraser University (1984). In fall 1984 I came to North Dakota State University where I am currently associate professor. I love international folk dancing and traveling with my daughters and sons Tammy, Ziva, Ilan and Ofer.

where F is an arbitrary analytic function in a domain of a complex plane. This example is a typical one, illustrating the use of the Laplace transform inversion-by-inspection for the construction of nonseparable solutions to some linear partial differential equations, and linear boundary value problems.

Let

$$\sum_{k=1}^{m} a_k(\mathbf{x}) \frac{\partial^{i_{1,k}+i_{2,k}+\cdots+i_{p,k}+j_k}}{\partial x_1^{i_{1,k}} \partial x_2^{i_{2,k}} \cdots \partial x_p^{i_{p,k}} \partial t^{j_k}} \psi(\mathbf{x}, t) = 0 \tag{4}$$

be a linear partial differential equation, called **actual equation**, in the $p+1$ real variables $\mathbf{x} = (x_1, x_2, \dots, x_p)$ and t, where $a_k(\mathbf{x})$, $1 \leqslant k \leqslant m$, are analytic in a domain of their variables. We associate the actual equation (4) with an **operational equation**, i.e., with the reduced linear partial differential equation

$$\sum_{k=1}^{m} s^{j_k} a_k(\mathbf{x}) \frac{\partial^{i_{1,k}+i_{2,k}+\cdots+i_{p,k}}}{\partial x_1^{i_{1,k}} \partial x_2^{i_{2,k}} \cdots \partial x_p^{i_{p,k}}} \phi(\mathbf{x}, s) = 0, \tag{5}$$

in the variables $\mathbf{x}$ and the real parameter s.

Equation (5) is an operational equation in the sense that the differential operator $\partial/\partial t$ of the actual equation (4) is treated in (5) as if it were an algebraic quantity s. A particular example of an actual equation and its associated operational equation is the Laplace equation (3) and its reduced equation (2).

Recalling the Laplace transform definition [**1**],

$$L\{ f(\mathbf{x}, t)\} = \int_0^{\infty} f(\mathbf{x}, t) e^{-st}\, dt,$$

where the Laplace transform operator L takes an **actual function** of $\mathbf{x}$ and t into an **operational function** of $\mathbf{x}$ and s, we see that the inverse Laplace transform L^{-1} of the integral $\int_0^\infty f(\mathbf{x}, t) e^{-st}\, dt$ is, by **inspection** and the definition itself,

$$L^{-1} \int_0^{\infty} f(\mathbf{x}, t) e^{-st}\, dt = f(\mathbf{x}, t). \tag{6}$$

It is well known in the operational calculus of the Laplace transform [**1**] that, under some general conditions, the inverse Laplace transform operator L^{-1} takes solutions of (5), called **operational solutions**, into solutions of (4), called **actual solutions.** These conditions involve the existence of some integrals and their interchangeability with differentiations. There is no need for us to consider these conditions on integrals, since all integrals involved will be eliminated in the process of our Laplace transform inversion. In this process we will reduce expressions into the form (6), enabling the Laplace transform inversion, L^{-1}, to be done by inspection. The reduction is achieved by means of a convenient change of variable formula that we present in equation (10) below.

For the presentation of a convenient change of variable formula, let us consider the indefinite complex integral

$$\int F(\tau, t^*(\mathbf{x}; \tau))\, d\tau, \tag{7}$$

in which the variable of integration τ appears in the argument of F explicitly, and implicitly through the function t^*. Both F and t^* are analytic functions of their variables in a domain of their spaces. Furthermore, we assume that $\partial t^*/\partial \tau \neq 0$ in the domain under consideration.

One can transform the variable of integration τ in (7) into a new variable of integration t by means of the substitution $t = t^*(\mathbf{x}; \tau)$ obtaining

$$\int F(\tau^*(\mathbf{x}; t), t)\, \dot{\tau}^*(\mathbf{x}; t)\, dt, \tag{8}$$

where $\tau^*(\mathbf{x}; t)$, replacing the old variable of integration τ, is a function of the new variable of

integration t and of $\mathbf{x}$, and where

$$\dot{\tau}^*(\mathbf{x}; t) = \frac{\partial \tau^*(\mathbf{x}; t)}{\partial t} \neq 0 \tag{9}$$

in the domain under consideration.

There is an obvious connection between the functions $t^*(\mathbf{x}; \tau)$ and $\tau^*(\mathbf{x}; t)$ in (7) and (8). These functions are inverse to one another in the sense that

$$t^*(\mathbf{x}; \tau^*(\mathbf{x}; t)) = t \quad \text{and} \quad \tau^*(\mathbf{x}; t^*(\mathbf{x}; \tau)) = \tau$$

are identities for $\mathbf{x}$, t and τ in some domains of their spaces.

The transformation from (7) to (8) is the well-known rule for changing a variable of integration, often called integration by substitution. If we omit the arguments of $\tau^*(\mathbf{x}; t)$ and $t^*(\mathbf{x}; \tau)$, the change of variable formula can thus be written as

$$\int F(\tau, t^*)\, d\tau = \int F(\tau^*, t)\, \dot{\tau}^*\, dt, \tag{10}$$

where t^* and τ^* are inverse to one another, and $\dot{\tau}^*$ is given by equation (9).

In the change of variable formula (10) the property of dependence of a variable is denoted by an asterisk. Since variables of integration are independent, the change of variable of integration is associated with the transition of the asterisk from one variable to another.

Let

$$F(\mathbf{x}; \tau)\, e^{-st^*(\mathbf{x}; \tau)} \tag{11}$$

be a **parametric** operational solution of an operational equation (5), F being analytic in a domain of its variables and parameter. The complex parameter in (11) is τ. The operational equation under consideration, equation (5), is **linear**, and **independent** of τ. Hence, a complex contour integral of (11) with respect to τ yields another operational solution to equation (5),

$$\int_\Gamma F(\mathbf{x}; \tau)\, e^{-st^*(\mathbf{x}; \tau)}\, d\tau. \tag{12}$$

Therefore, an actual solution, i.e., a solution to the actual equation (4), is given by the inverse Laplace transform of (12),

$$L^{-1}\left\{ \int_\Gamma F(\mathbf{x}; \tau)\, e^{-st^*(\mathbf{x}; \tau)}\, d\tau \right\}, \tag{13}$$

where Γ is any complex path of integration in the complex τ-plane. In particular, and as an example, if (11) satisfies the reduced Laplace equation (2), then (13) satisfies Laplace equation (3). Here we do not pay attention to the existence or non-existence of the integral (12); the reason will be clarified in the paragraph preceding equation (26). We are now going to manipulate the integral in (13) into a form which enables the Laplace transform inversion, L^{-1}, to be done by inspection, as in equation (6).

Employing the change of variable formula (10), we manipulate the expression (13) into

$$L^{-1}\left\{ \int_{\Gamma'} F(\mathbf{x}; \tau^*)\, \dot{\tau}^* e^{-st}\, dt \right\}, \tag{14}$$

where Γ', replacing Γ, is the new path of integration owing to the change of the variable of integration.

The path of integration Γ in (12) has not yet been specified. If we now specify it to be such that its map Γ' in (14) is the positive ray, $0 \leqslant t < \infty$, then (14) becomes

$$L^{-1}\left\{ \int_0^\infty F(\mathbf{x}; \tau^*)\, \dot{\tau}^* e^{-st}\, dt \right\}, \tag{15}$$

which is equal, by inspection, to

$$F(\mathbf{x}; \tau^*)\dot{\tau}^*. \tag{16}$$

The transformation

$$F(\mathbf{x}; \tau)e^{-st^*(\mathbf{x};\, \tau)} \to F(\mathbf{x}; \tau^*)\dot{\tau}^* \tag{17}$$

from the parametric operational solution (11) into the actual solution (16) involves a parametric integration, (12), destined to play against the Laplace transform inversion, (13), to the point of their mutual annihilation, (15). An illustrative example of transformation (17) follows.

The expression

$$F(\tau)e^{-st^*}, \tag{18}$$

where

$$t^*(x, y; \tau) = x\sinh\tau - iy\cosh\tau, \quad -\infty < x, y < \infty, \tag{19}$$

satisfies the reduced Laplace equation (2) for any analytic function F of τ. Since the parameter τ does not appear in (2), any function of τ can be regarded as a constant with respect to (2) and, hence, the arbitrariness of F in (18) is trivial. It is, however, not trivial in the actual solution $F(\tau^*)\dot{\tau}^*$ which we develop in equation (23) below. Regarding $F(\tau)$, $\sinh\tau$ and $\cosh\tau$ as constants, the solutions (18) and (1) to the reduced Laplace equation (2) are obviously equivalent.

By the Laplace transform inversion procedure, described in equations (11)–(16) and summarized in transformation (17), the parametric operational solution (18) of the operational equation (2) yields an actual solution, $F(\tau^*)\dot{\tau}^*$, to the actual equation (3). Here the function $\tau^*(x, y; t)$ in $F(\tau^*)\dot{\tau}^*$ is the inverse of $t^*(x, y; \tau)$ of (19). Hence, from equation (19), τ^* is defined implicitly by

$$t = x\sinh\tau^* - iy\cosh\tau^* = r\sinh(\tau^* - i\theta), \quad r^2 = x^2 + y^2, \quad \theta = \tan^{-1}\left(\frac{y}{x}\right), \tag{20}$$

implying that

$$\tau^*(x, y; t) = \sinh^{-1}\left(\frac{t}{r}\right) + i\theta, \qquad r \neq 0, \tag{21}$$

the derivative with respect to t of which is

$$\dot{\tau}^* = \frac{1}{\sqrt{r^2 + t^2}}. \tag{22}$$

We may note that equation (20) is obtained from equation (19) by removing the asterisk from t and imposing it on τ, thus indicating that the property of dependence of a variable is transferred from t in (19) to τ in (20).

As expected, the solution $F(\tau^*)\dot{\tau}^*$ which, by means of equations (21) and (22), can be written as

$$F(\tau^*)\dot{\tau}^* = \frac{1}{\sqrt{r^2 + t^2}}F\left(\sinh^{-1}\left(\frac{t}{r}\right) + i\theta\right), \qquad r \neq 0, \tag{23}$$

satisfies Laplace equation (3) in a domain of the (x, y, t)-space for an **arbitrary** analytic function F defined in a domain of a complex plane.

In this example the Laplace transform inversion by inspection, (17), was employed to transform the parametric operational solution (18) of the operational equation (2) into the actual solution (23) of the actual equation (3):

$$F(\tau)e^{-s(x\sinh\tau - iy\cosh\tau)} \to \frac{1}{\sqrt{r^2 + t^2}}F\left(\sinh^{-1}\left(\frac{t}{r}\right) + i\theta\right). \tag{24}$$

The particular transformation (24) illustrates the usefulness of the general transformation (17); it takes a simple operational solution into an advanced actual solution.

The operational solution $F(\tau)e^{-st^*}$, (18), is equivalent to the operational solution (1). It is simpler than the resulting actual solution, (23), in the sense that

(a) the two variables x and y in (18) are separable while the three variables x, y and t in (23) are not separable; and

(b) the arbitrariness of the function F in (18) is trivial, since it is a function of a parameter, while its arbitrariness in (23) is not trivial, since it is a function of variables.

The solution (23) of Laplace equation (3) is known in the literature, but in similar ways unknown solutions to linear partial differential equations and linear boundary value problems can be discovered. An example is the solution (25) of the n-dimensional wave equation, discovered in [2] using the techniques of this article:

$$\omega = \left(R_n + \sum_{i=1}^{n} \alpha_i x_i\right)^{(3-n)/2} \frac{1}{R_n} F(R_n - ct), \tag{25}$$

where

$$R_n^2 = \sum_{i=1}^{n} x_i^2, \quad \sum_{i=1}^{n} \alpha_i^2 = 1, \quad \alpha_i = \text{constant}, \quad n = 1,2,3,\ldots,$$

is a solution for $\displaystyle\sum_{i=1}^{n} \frac{\partial^2 \omega}{\partial x_i^2} = \frac{1}{c^2}\frac{\partial^2 \omega}{\partial t^2}$.

The solution (25) seems to be novel for $n > 3$, indicating the known fact that only in our three-dimensional world is high-fidelity signal transmission from point sources possible. Only the case of $n = 3$ in (25) admits a representation of distortionless waves emitted from a point source; see Courant [3].

The conditions under which the Laplace transform inversion by inspection, (17), is justified, are beyond those necessary for the existence of the integrals in (12), (14) and (15). One can thus define an operator U that sends the operational solution (11) to the actual one (16),

$$U\{F(\mathbf{x};\tau)e^{-st^*}\} = F(\mathbf{x};\tau^*)\dot{\tau}^*, \tag{26}$$

and on revealing its properties a new operational calculus emerges [2].

In the new operational calculus, the definition of the operator U, referred to as the **differential transform** is extended to

$$U\{s^n F(x;t)e^{-st^*}\} = \partial_t^n\{F(x;\tau^*)\dot{\tau}^*\}, \tag{27}$$

where ∂_t^n denotes partial differentiation of order n with respect to t, and n is an integer. If $n < 0$, ∂_t^n means integration of order $-n$ with respect to t [4].

The definition in (27), yielding $Us = \partial_t U$, is obviously motivated by the property $L^{-1}s = \partial_t L^{-1}$ of the inverse Laplace transform operator L^{-1}. While this relationship for L^{-1} is valid under some well-known restrictions on the functions involved in the operation of L^{-1}, the validity of the relationship $Us = \partial_t U$ is a result of a definition. The worth of that definition rests on the resulting commutativity relationship $U\partial_{x_k} = \partial_{x_k}U$, $1 \leqslant k \leqslant p$, enjoyed by the differential transform U of (27). A proof of this is given in [5].

The differential transform U shares its commutativity property with the inverse Laplace transform operator L^{-1}, $L^{-1}\partial_{x_k} = \partial_{x_k}L^{-1}$, $1 \leqslant k \leqslant p$. The validity of this last commutativity depends on the behavior of the functions involved, e.g., their behavior must be good enough to allow integrations and differentiations to be interchanged. The differential transform, being an integral-free transform, is obviously not subject to this sort of restrictions.

These various "commutativity" properties of U are those upon which the operational calculus of the differential transform is based. It is due to these properties that the differential transform is grouped with integral transforms; see Sneddon [6].

Equation (26) presents a **differential transform representation** representing $F(\mathbf{x}, \tau^*)\dot{\tau}^*$ as the differential transform of $F(\mathbf{x}, \tau)e^{-st^*}$. Accordingly, the transformation (24) can be written as the differential transform representation

$$U\{F(\tau)e^{-st^*}\} = \frac{1}{\sqrt{r^2+t^2}} F\left(\sinh^{-1}\left(\frac{t}{r}\right) + i\theta\right), \tag{28}$$

where

$$t^*(x, y; \tau) = x \sinh \tau - iy \cosh \tau$$

and

$$r^2 = x^2 + y^2, \qquad \theta = \tan^{-1}(y/x),$$

and where F is an arbitrary analytic function of a complex variable in a domain. Equation (28) represents an advanced solution of Laplace equation (3) as the differential transform of a simpler parametric solution of the simpler, reduced Laplace equation (2).

Another differential transform representation illustrating equation (26) is given by the equation

$$U\{F(x\cosh\xi - iy\sinh\xi - t, \xi)e^{-sz^*}\} = \frac{1}{R} F\left(R - t, \sinh^{-1}\left(\frac{z}{r}\right) + i\theta\right), \tag{29}$$

where

$$z^*(x, y; \xi) = x\sinh\xi - iy\cosh\xi$$

and

$$R^2 = r^2 + z^2,$$

and where F is an arbitrary analytic function of two complex variables in a domain. Equation (29) represents an advanced solution of the **wave equation,**

$$\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2} = \frac{\partial^2\psi}{\partial t^2},$$

as the differential transform of a simple parametric solution of the **reduced wave equation,**

$$\frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + s^2\phi = \frac{\partial^2\phi}{\partial t^2}.$$

Differential transform representations have a role similar to that of integral representations. In applications to linear boundary value problems, the differential transform represents advanced solutions as the differential transform of simple parametric solutions to corresponding boundary value problems. Examples are provided in [7], [8], [9].

References

1. P. K. F. Kuhfittig, Introduction to the Laplace Transform, Plenum Press, New York, 1978.

2. A. Ungar, A new operational calculus method, Internat. J. Engrg. Sci., 18 (1980) 43–59.

3. R. Courant, Hyperbolic partial differential equations and applications, in Modern Mathematics for the Engineer, edited by E. F. Beckenbach, McGraw-Hill, New York, 1956.

4. A. Ungar, The differential transform technique for solving problems of wave propagation, in Modern Problems in Elastic Wave Propagation, edited by J. Miklowitz and J. D. Achenbach, Wiley, 1978, pp. 83–102.

5. A. Ungar, An operator related to the inverse Laplace transform, SIAM J. Math. Anal., 5 (1974) 367–375.

6. I. N. Sneddon, Recent applications of integral transforms in the linear theory of elasticity, in Trends and Applications of Pure Mathematics to Mechanics, Vol. II, edited by H. Zorski, London, 1977, pp. 307–325.

7. A. Ungar and N. I. Robinson, The application of the differential transform technique to point source problems of elasticity, Internat. J. Engrg. Sci., 15 (1977) 157–170.

8. A. Ungar, A note on Ranger's line rotelet near a plane, Int. J. Comp. Math. Appl., 8 (1982) 95–102.

9. A. Ungar and C. H. Murrell, The differential transform and its application to an electrostatics image problem, Internat. J. Comput. Math. Appl., 11 (1985) 565–572.

His last name is frequently last but never least. (See p. 801.)

A similar problem is

Q4. Does there exist a closed set S in 3-dimensional space such that $\lambda_2(S)$ is minimal and $B_3 \to S$, where B_3 is the unit ball?

In 1974, R. Laver found that for every $\varepsilon > 0$ there exists a set S with $B_3 \to S$ and $\lambda_2(S) < 2\pi + \pi^2/2 + \varepsilon$. This set S (see Fig. 11) consists of the lower hemisphere (of area 2π) plus vertical rings standing on the surface of the sphere. The first ring stands on the equator, the second ring stands on the circle formed by the intersection of the plane of the top of the first ring and the sphere, etc. The height of the last ring extends to the north pole. Since

$$\int_0^{\pi/2} 2\pi \cos\alpha \, d(\sin\alpha) = \frac{\pi^2}{2},$$

it is clear that if the consecutive rings are narrow enough, then their joint area is less than $\pi^2/2 + \varepsilon$. It is also clear that every line that intersects the sphere intersects S. We do not know if there exists an S with $B_3 \to S$ and $\lambda_2(S) = 2\pi + \pi^2/2$.

Editorial note. The authors have since shown that the Steiner tree of a triangle T is the shortest connected set which meets all the lines which meet T.

The problem has already appeared [**2, 3, 5, 7**] in various forms: a hunter lost in a dense forest who knows he is within a mile of a straight boundary; a swimmer at sea in a thick fog who knows she is within a mile of a straight shoreline. Each has zero visibility, but can do dead reckoning navigation. The title of Croft's paper [**2**] doesn't immediately suggest a connexion, but see his section 3; he gives Ogilvy [**7**] as his source. Eggleston [**3**] solves the question originally asked by Croft, and also by the authors and others:

Q0. What is the shortest (in terms of λ_1) connected closed set such that $B_2 \to S$?

by maximizing the radius of the disc, rather than by minimizing the length of the connected set. The answer is as shown in Fig. 2; a special case was earlier considered by Joris [**5**]. Moran considers a related problem and mentions others at the end of his paper [**6**].

References

1. F. R. K. Chung and R. L. Graham, Steiner trees for ladders, Ann. Discrete Math., 2 (1978) 173–200; MR 58#315.

2. H. T. Croft, Curves intersecting certain sets of great circles on the sphere, J. London Math. Soc. (2) 1(1969) 461–469; MR 40#865.

3. H. G. Eggleston, The maximal inradius of the convex cover of a plane connected set of given length, Proc. London Math. Soc. (3) 45(1982) 456–478; MR 84e:52004.

4. Vance Faber, Jan Mycielski, and Paul Pedersen, On the shortest curve which meets all the lines which meet a circle, Ann. Polon. Math., 44(1984) 249–266.

5. H. Joris, Le chasseur perdu dans la forêt: un problème de géométrie plane, Elem. Math., 35(1980) 1–14; MR 81d:52001.

6. P. A. P. Moran, On a problem of S. Ulam, J. London Math. Soc., 21(1946) 175–179; MR 8, 597n.

7. C. Stanley Ogilvy, Tomorrow's Math, Unsolved Problems for the Amateur, 2nd ed., Oxford University Press, 1972 (1st ed. 1962, pp. 23–24).

8. R. Spira, The isogonic and Erdős-Mordell points of a simplex, this MONTHLY, 78(1971) 856–864; MR 44#5861.

ANSWER TO PHOTO ON PAGE 792

Antoni Zygmund.

PROGRESS REPORTS

EDITED BY THOMAS BANCHOFF AND RICHARD MILLMAN

It is easy to be too busy to pay attention to what anyone else is doing, but not good. All of us should know, and want to know, what has been discovered since our formal education ended, but new words, and relations between them, are growing too fast to keep up. It is possible for a person to learn of the title of a recent work and of the key words used in it and still not have the faintest idea of what the subject is.

Progress Reports is to be an almost periodic column intended to increase everyone's mathematical information about what others have been up to. Each column will report one step forward in the mathematics of our time. The purpose is to inform, more than to instruct: what is the name of the subject, what are some of the words it uses, what is a typical question, what is the answer, who found it. The emphasis will be on concrete questions and answers (theorems), and not on general contexts and techniques (theories). References will be kept minimal: usually they will include only one of the earliest papers in which the answer appears and a more recent exposition of the discovery, whenever one is easily available.

Everyone is invited to nominate subjects to be reported on and authors to prepare the reports. The ground rules are that the principal theorem should be old enough to have been published in the usual sense of that word (and not just circulated by word of mouth or in preprints); it should be of interest to more than just a few specialists; and it should be new enough to have an effect on the mathematical life of the present and near future. In practice most reports will probably be on progress achieved somewhere between 5 and 15 years ago.

POINTS OF FINITE ORDER ON ELLIPTIC CURVES

JOSEPH H. SILVERMAN
Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139

The problem of solving polynomial equations in integers or rational numbers dates back to antiquity. Quadratic equations are by this time fairly well understood, but already cubic equations in two variables, so called "elliptic curves," present many challenges beyond current resources to solve. We will look at equations

$$(*) \qquad y^2 = x^3 + Ax + B,$$

where $A, B \in \mathbb{Q}$ are rational numbers and the discriminant $\Delta = 4A^3 + 27B^2$ is non-zero, and examine the solutions in rational numbers $x, y \in \mathbb{Q}$.

Mordell in 1922 had described the nature of the solution set, but he left open the question of the number of points of finite order. Progress in the late 30's gave a criterion which such points had to satisfy, but it was not known whether the number of such points could be arbitrarily large. Then a beautiful result by Mazur in 1977 showed that there could never be more than fifteen such points!

If $P = (x_1, y_1)$ and $Q = (x_2, y_2)$ are solutions to $(*)$, we can usually produce a new solution $R = (x_3, y_3)$ as follows. (See Fig. 1.) Draw the line L through P and Q. (If $P = Q$, we take L to be the tangent line to the curve at the point P.) Then, since $(*)$ has degree 3, the line L will intersect $(*)$ in a third point $R = (x_3, y_3)$ (unless the line L happens to be vertical; more about that later).
Further, if P and Q have coordinates in $\mathbb{Q}$, then so does R, a fact pointed out by Poincaré around 1900. (You can check this.) We want to define the "sum" of P and Q, and it might seem natural to take it to be R, but it turns out to be better to reflect through the x-axis and define

$$P + Q = (x_3, -y_3).$$

This addition is commutative (easily checked) and associative (not so easy). We also define an inverse by $-P = (x_1, -y_1)$. To round matters out, if the line through P and Q is vertical, we set

$$P + Q = \infty,$$

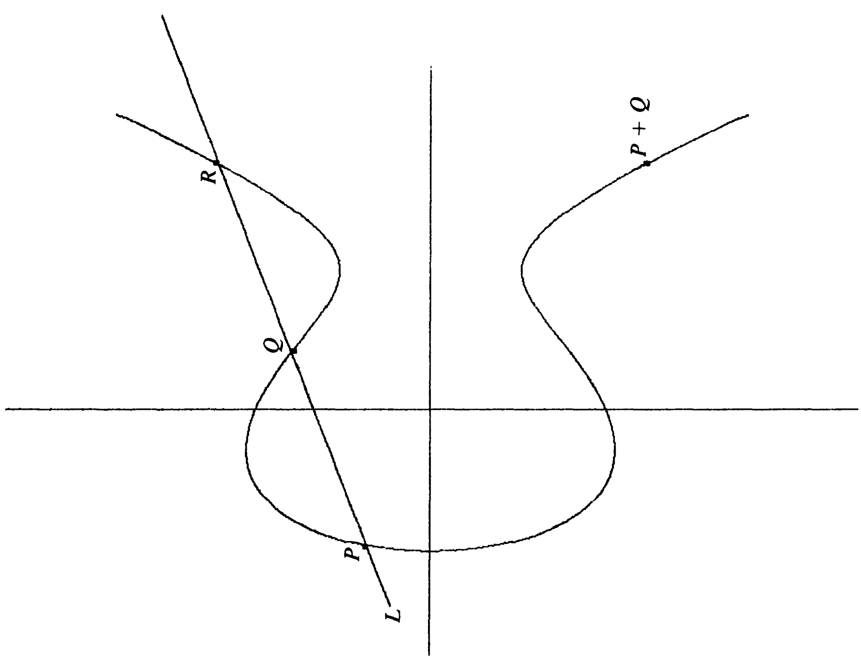


FIG. 1

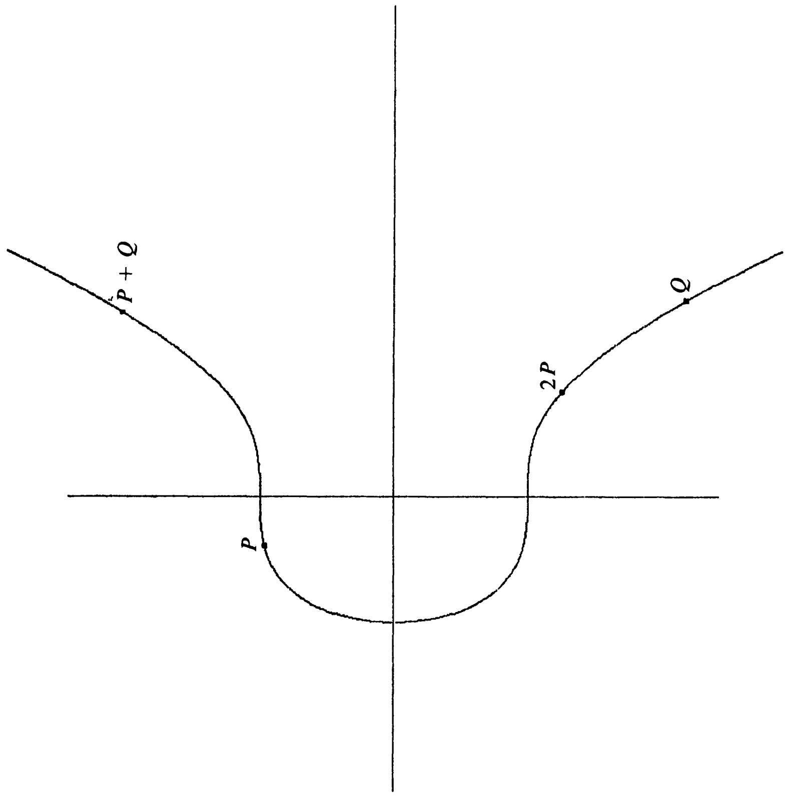


FIG. 2

where ∞ denotes an "extra" solution to $(*)$. (The solutions of $(*)$ and the point ∞ then form what is known as an abelian group.) You can check the following examples. (See Fig. 2.)

EXAMPLE 1. $y^2 = x^3 + 17$.
$P = (-1, 4)$, $Q = (4, -9)$, $2P = (137/64, -2651/512)$, $P + Q = (94/25, 1047/125)$.

EXAMPLE 2. $y^2 = x^3 - 372x + 2761$.
$P = (2, 45)$, $2P = (12, -5)$, $3P = (11, 0)$, $4P = (12, 5)$, $5P = (2, -45)$, $6P = \infty$.

The fundamental theorem of Mordell says that to find all rational solutions to $(*)$, one can start with a finite set and use the above addition rules to find all the others.

THEOREM (Mordell [3]). *There is a finite set* $\{P_1, \ldots, P_r\}$ *of rational solutions to* $(*)$ *so that every rational solution P equals*

$$n_1P_1 + n_2P_2 + \cdots + n_rP_r$$

for some integers $n_1, \ldots, n_r$. (In other words, the group of rational solutions is finitely generated.)

Now sometimes if you start with a solution P and form its multiples $2P, 3P, \ldots,$ the list extends indefinitely. This happens in Example 1, as we will see below. On the other hand, as in Example 2, sometimes the list terminates with $nP = \infty$. If this latter occurs, we say that P has *finite order*, and its *order* is equal to the smallest $n \geq 1$ for which $nP = \infty$. The following theorem allows one to actually find all of the points of finite order, of which a priori there are only finitely many by Mordell's theorem.

THEOREM (Lutz-Nagell [1], [4]). *Assume the equation* $(*)$ *is written so that A and B are integers, and let* $P = (x, y)$ *be a rational solution of finite order. Then x and y are integers, and either*

(i) $y = 0$, *in which case* $2P = \infty$;

or

(ii) y^2 *divides* $\Delta = 4A^3 + 27B^2$.

Thus in Example 2 we have $\Delta = -3^6 5^3$, which is divisible by $y^2 = 3^4 5^2$; while in Example 1, since $2P$ does not have integer coordinates, it follows that P is not of finite order.

Now for any given equation $(*)$ we can use the Lutz-Nagell theorem to find all of the rational points of finite order; but if A and B are large, one might suppose that there could be many such points. In fact, this is not the case, a deep and beautiful result recently proven by Barry Mazur.

THEOREM (Mazur [2]). *The equation* $(*)$ *has at most* 15 *rational points of finite order. If it has a rational point of order* n, *then* $2 \leq n \leq 10$ *or* $n = 12$.

Further, for each $n \leq 10$ and $n = 12$, one can find an equation $(*)$ with a rational point of order n, so Mazur's theorem actually gives the entire story for rational points of finite order on elliptic curves.

Currently the main direction for further exploration is to generalize the results to arbitrary number fields; but even more important, the whole area is intimately connected with the theory of modular curves and modular functions, which are tied up with a number of interesting conjectures and are presently the subject of much study.

(Computer generated illustrations produced by Thomas Banchoff and Paul Strauss at Brown University.)

References

1. E. Lutz, Sur l'équation $y^2 = x^3 - Ax - B$ dans les corps p-adiques, J. Reine Angew. Math., 177 (1937) 237–247.

2. B. Mazur, Modular curves and the Eisenstein ideal, Inst. Hautes Études Sci. Publ. Math., 47 (1977) 33–186.

3. L. J. Mordell, On the rational solutions of the indeterminate equation of the third and fourth degrees, Proc. Cambridge Philos. Soc., 21 (1922) 179–192.

4. T. Nagell, Solution de quelques problèmes dans la théorie arithmétique des cubiques planes du premier genre, Skr. Norske Vid.-Akad. Oslo I (1935), Nr. 1.

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

THE SHORTEST CURVE THAT MEETS ALL THE LINES THAT MEET A CONVEX BODY

V. FABER

Los Alamos National Laboratory, Los Alamos, NM 87545

J. MYCIELSKI

Department of Mathematics, University of Colorado, Boulder, CO 80309-0426

A telephone company, while repairing buried cable, has discovered that although the cable is buried 1 m deep, often the cable is not directly under the marker that is supposed to be erected above it. They do know that the cable is always within 2 m of the marker in the horizontal plane. To ensure finding the cable, even when its direction is unknown, the repairmen dig a 1-m-deep trench in a circle of radius 2 m about the marker (see Fig. 1). In 1974, M. Magidor showed us that Fig. 2 gives a more efficient way of finding the cable, and he asked for the length of the shortest trench that will find the cable (assuming that the cable is straight and *does* lie within 2 m of the marker). We demonstrated in [4] that for a continuous trench this method is optimal.

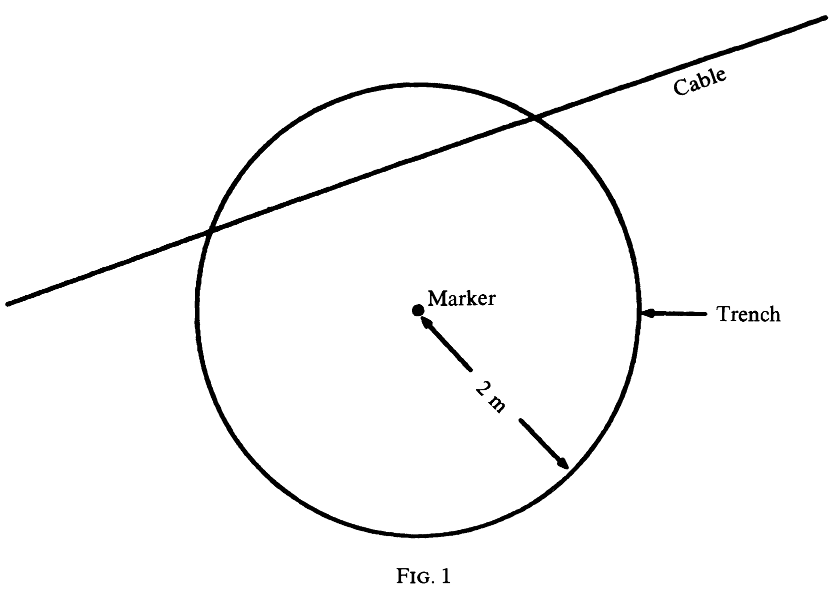


FIG. 1

What is the shortest curve that meets all the lines that meet a given circle? The answer may depend on our meaning of the words "curve" and "length," as we shall see below. Let $R \to S$ mean that R and S are sets in the plane (or more generally, n-dimensional Euclidean space) and every straight line meeting the set R meets the set S. Some short curves S such that $C \to S$ are shown in Figs. 3–9 for various convex sets C (lengths shown are percentages of the length of the boundary, where "length" will be defined below). Figs. 3–6 represent regular polygons. The

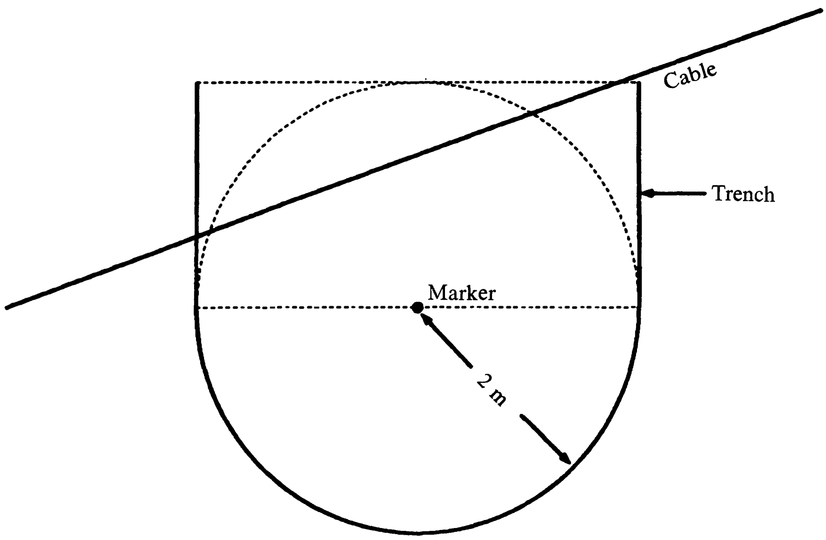
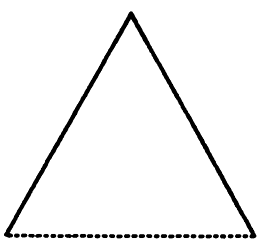
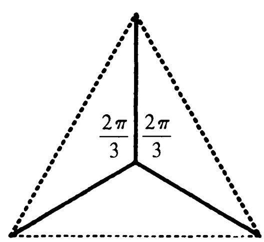


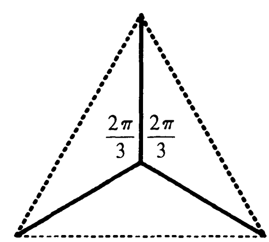
FIG. 2



(a) $\frac{2}{3} = .666666666\ldots$

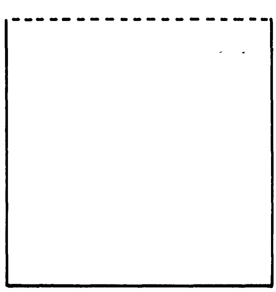


(b) $\frac{\sqrt{3}}{3} = .577350269\ldots$

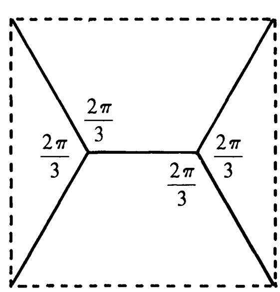


(c) $\frac{\sqrt{3}}{3} = .577350269\ldots$

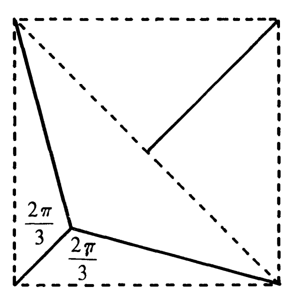
FIG. 3



(a) $\frac{3}{4} = .75$



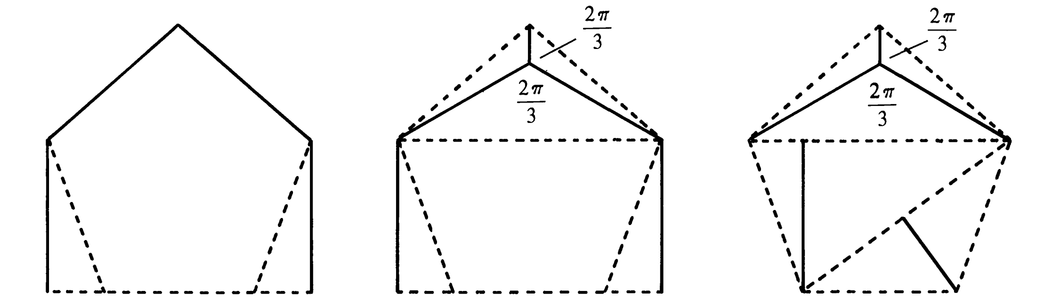
(b) $\frac{1+\sqrt{3}}{4} = .683012702\ldots$



(c) $.659739609\ldots$

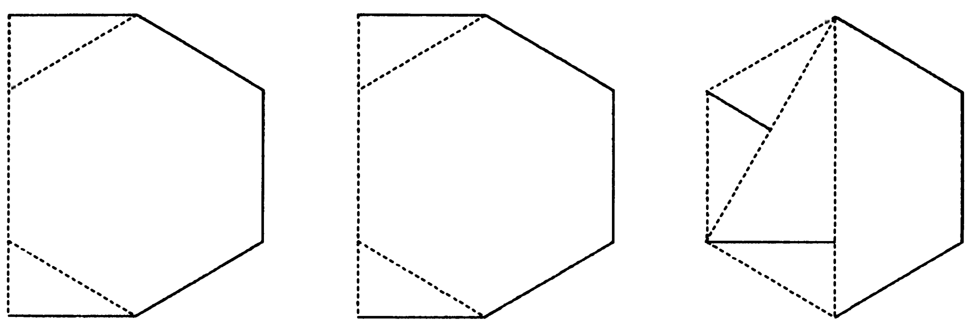
FIG. 4

figures (a) illustrate the shortest known arcs, the figures (b) illustrate the shortest known connected sets, and the figures (c) illustrate the shortest known closed sets. Note that in Figs. 3(b) and (c), 4(b), and 7(b), the shortest known sets join all the vertices of the given polygon—the so called *Steiner span* of the vertices (see **[1]** and **[8]**). (The Steiner span seems to be different from the shortest connected set meeting all the lines that meet a given polygon when that polygon has many sides.)



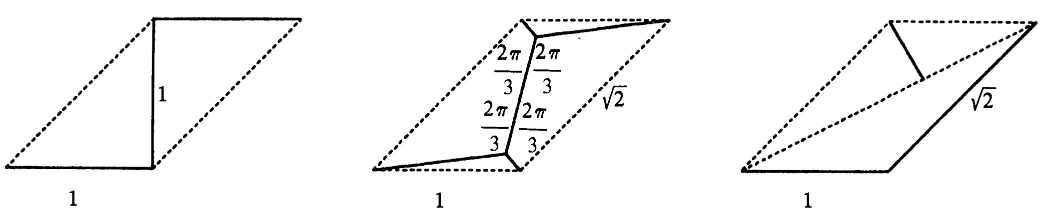
(a) .780422607... (b) .778231365... (c) .705577112

FIG. 5



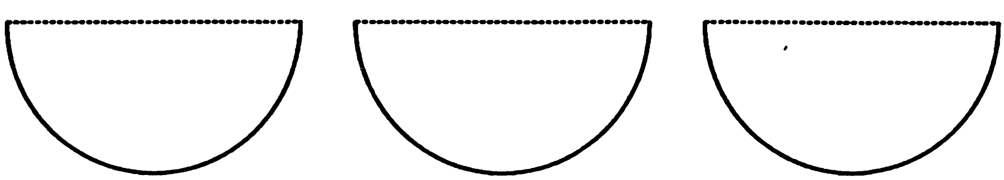
(a) $\frac{3+\sqrt{3}}{6} = .788675135\ldots$ (b) .788675135... (c) $\frac{7+\sqrt{3}}{12} = .727670901\ldots$

FIG. 6



(a) $\frac{3}{2+2\sqrt{2}} = .621320344\ldots$ (b) .602538433... (c) .592620968...

FIG. 7



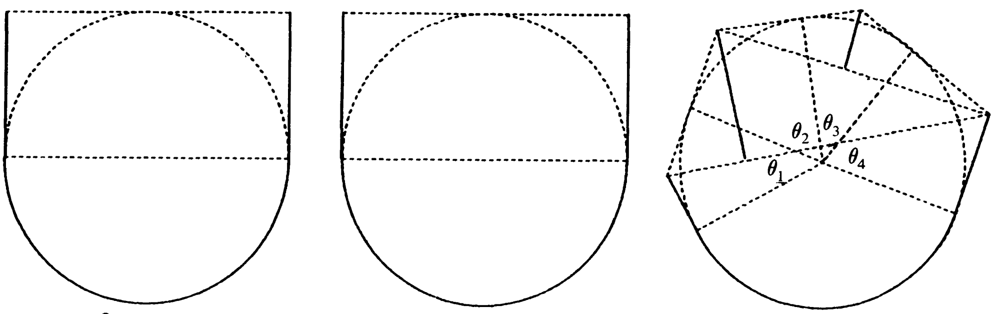
(a) $\frac{\pi}{\pi+2} = .611015470\ldots$ (b) .611015470... (c) .611015470...

FIG. 8

If we restrict our attention to *paths*, that is, continuous functions $f(t) = (x(t), y(t))$ from the unit interval I into the plane, the natural length to consider is *path length*:

$$\lim_{n\to\infty} \sum_{k=0}^{n-1} \left\{ \left(x((k+1)/n) - x(k/n) \right)^2 + \left(y((k+1)/n) - y(k/n) \right)^2 \right\}^{1/2}.$$

A path f is called an *arc* if f is one-to-one. It was shown in [4] that the shortest path that meets all the lines that meet the unit circle is an arc and has length $\pi + 2$ [Fig. 9(a)].



(a) $\dfrac{\pi + 2}{2\pi} = .818309886\ldots$ (b) $.818309886\ldots$ (c) $.763926466\ldots$

$\theta_1 = .96 \quad \theta_2 = 1.04$

$\theta_3 = .7 \quad \theta_4 = 1.2$

FIG. 9

A natural generalization is the *n-arc*; that is, a sequence $P_n = (f_1, f_2, \ldots, f_n)$ where each f_i is an arc. If P_n is an n-arc, its *arc length* is the sum of the arc lengths of the f_i's.

We can measure the length of any Borel set in n-dimensional Euclidean space by means of the *one-dimensional Hausdorff measure* λ_1. The *α-dimensional Hausdorff outer measure* (where α is a positive real number) of any subset S of a metric space is

$$\lambda_\alpha(S) = \lim_{\delta \to 0}\left(\inf\left\{\sum_{i=1}^{\infty} (\operatorname{diam} E_i)^\alpha \,\Big|\, \bigcup_{i=1}^{\infty} E_i = S \text{ and } \operatorname{diam} E_i \leqslant \delta \text{ for all } i\right\}\right).$$

For example, if C is the Cantor set, then $\lambda_1(C) = 0$ but $\lambda_1(C^2) = \infty$.

Little is known about the relationship $R \to S$. In [3] and [4], it is shown that: (1) if R is closed and bounded, then there is a shortest (in terms of λ_1) closed connected set S such that $R \to S$ and its length is $\pi + 2$ (see Fig. 9); (2) if R is convex and ∂R is its boundary, then any closed set S such that $R \to S$ satisfies $\lambda_1(S) \geqslant \frac{1}{2}\lambda_1(\partial R)$; (3) if R is a compact convex set in the plane, then the shortest path f such that $R \to fI$ is one-to-one. Now we shall prove a theorem that answers a question stated in Section 2 of [4].

THEOREM. *If R is a compact set in the plane, then there exists a closed set S satisfying $R \to S$ and of minimal length (in terms of λ_1) among those with at most n connected components.*

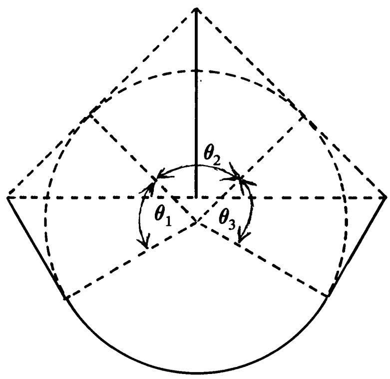
Proof. Let λ_0 be the infimum of the lengths $\lambda_1(S)$ for all closed S of at most n components with $R \to S$. Let S_i have at most n components and satisfy $R \to S_i$ and $\lim \lambda_1(S_i) = \lambda_0$. By choosing an appropriate subsequence of the S_i, we can assume that $S_{i,1}, \ldots, S_{i,k}$ are components of S_i whose distance from R is bounded as $i \to \infty$, while $S_{i,k+1}, \ldots, S_{i,n}$ are the components whose distance from R is unbounded. By again choosing an appropriate subsequence of the S_i, we can assume that, for $j \leqslant k$, S_{ij} converges in Hausdorff distance to certain closed sets S_j^* (as $i \to \infty$) and that, for $j > k$, the slopes of the lines meeting S_{ij} and R converge to some angles α_j. Let $S^* = S_1^* \cup \cdots \cup S_k^*$. Thus the only lines meeting R but missing S^* would have one of the slopes $\alpha_{k+1}, \ldots, \alpha_n$. But, since S^* is compact, if there exists any line L meeting R and missing S^*, then for any $p \in L$, any L' containing p and such that the angle between L and L' is small enough also misses S^*. Thus a finite number of exceptional slopes is impossible, and $R \to S^*$ follows.

But since the S_{ij} are connected and the sequence S_{ij} converges in Hausdorff distance to S_j^*, it follows that the S_j^* are connected and

$$\lambda_1(S_j^*) \leqslant \liminf_{i \to \infty} \lambda_1(S_{ij})$$

(for a simple proof and references to related inequalities, see for example [**4**, Theorem 3]). Hence $\lambda_1(S^*) \leqslant \lambda_0$, and, since $R \to S^*$, $\lambda_1(S^*) = \lambda_0$. This concludes the proof of the theorem.

Now we list some open questions. Let B_2 be the unit disk.



$\theta_1 = \theta_3 = 1.28652\ldots$

$\theta_2 = 1.19106\ldots$

FIG. 10

Q1. Does the shortest n-arc P_n such that $B_2 \to P_n I$, have exactly n components? (A short 2-arc is shown in Fig. 10 and the shortest we know is a 3-arc shown in Fig. 9(c).)

Q2. Does there exist a shortest closed set S in the plane such that $B_2 \to S$?

Q3. For the regular polygon C_n with n sides, what is the shortest connected closed set S such that $C_n \to S$?

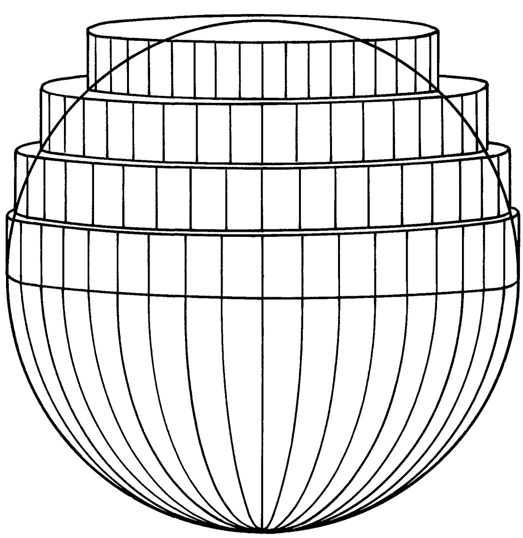


FIG. 11

A similar problem is

Q4. Does there exist a closed set S in 3-dimensional space such that $\lambda_2(S)$ is minimal and $B_3 \to S$, where B_3 is the unit ball?

In 1974, R. Laver found that for every $\varepsilon > 0$ there exists a set S with $B_3 \to S$ and $\lambda_2(S) < 2\pi + \pi^2/2 + \varepsilon$. This set S (see Fig. 11) consists of the lower hemisphere (of area 2π) plus vertical rings standing on the surface of the sphere. The first ring stands on the equator, the second ring stands on the circle formed by the intersection of the plane of the top of the first ring and the sphere, etc. The height of the last ring extends to the north pole. Since

$$\int_0^{\pi/2} 2\pi \cos\alpha \, d(\sin\alpha) = \frac{\pi^2}{2},$$

it is clear that if the consecutive rings are narrow enough, then their joint area is less than $\pi^2/2 + \varepsilon$. It is also clear that every line that intersects the sphere intersects S. We do not know if there exists an S with $B_3 \to S$ and $\lambda_2(S) = 2\pi + \pi^2/2$.

Editorial note. The authors have since shown that the Steiner tree of a triangle T is the shortest connected set which meets all the lines which meet T.

The problem has already appeared [**2**, **3**, **5**, **7**] in various forms: a hunter lost in a dense forest who knows he is within a mile of a straight boundary; a swimmer at sea in a thick fog who knows she is within a mile of a straight shoreline. Each has zero visibility, but can do dead reckoning navigation. The title of Croft's paper [**2**] doesn't immediately suggest a connexion, but see his section 3; he gives Ogilvy [**7**] as his source. Eggleston [**3**] solves the question originally asked by Croft, and also by the authors and others:

Q0. What is the shortest (in terms of λ_1) connected closed set such that $B_2 \to S$?

by maximizing the radius of the disc, rather than by minimizing the length of the connected set. The answer is as shown in Fig. 2; a special case was earlier considered by Joris [**5**]. Moran considers a related problem and mentions others at the end of his paper [**6**].

References

1. F. R. K. Chung and R. L. Graham, Steiner trees for ladders, Ann. Discrete Math., 2 (1978) 173–200; MR 58#315.

2. H. T. Croft, Curves intersecting certain sets of great circles on the sphere, J. London Math. Soc. (2) 1(1969) 461–469; MR 40#865.

3. H. G. Eggleston, The maximal inradius of the convex cover of a plane connected set of given length, Proc. London Math. Soc. (3) 45(1982) 456–478; MR 84e:52004.

4. Vance Faber, Jan Mycielski, and Paul Pedersen, On the shortest curve which meets all the lines which meet a circle, Ann. Polon. Math., 44(1984) 249–266.

5. H. Joris, Le chasseur perdu dans la forêt: un problème de géométrie plane, Elem. Math., 35(1980) 1–14; MR 81d:52001.

6. P. A. P. Moran, On a problem of S. Ulam, J. London Math. Soc., 21(1946) 175–179; MR 8, 597n.

7. C. Stanley Ogilvy, Tomorrow's Math, Unsolved Problems for the Amateur, 2nd ed., Oxford University Press, 1972 (1st ed. 1962, pp. 23–24).

8. R. Spira, The isogonic and Erdős-Mordell points of a simplex, this MONTHLY, 78(1971) 856–864; MR 44#5861.

ANSWER TO PHOTO ON PAGE 792

Antoni Zygmund.

NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.

For instructions about submitting Notes for publication in this department see the inside front cover.

FACTORING LARGE NUMBERS ON A POCKET CALCULATOR

W. D. BLAIR

Department of Mathematics, Northern Illinois University, DeKalb, IL 60115

C. B. LACAMPAGNE

Department of Mathematics, University of Michigan, Flint, MI 48503

J. L. SELFRIDGE

Department of Mathematics, Northern Illinois University, DeKalb, IL 60115

Factoring large numbers has long intrigued both amateur and serious number theorists, and factoring has been given increased attention with recent applications to cryptography. We present algorithms for factoring which can be featured in a first course in number theory and which form an attractive path to understanding many important concepts such as greatest common divisor, Fermat's little theorem, quadratic reciprocity calculations, Lucas sequences, etc.

In addition to a short discussion and a step-by-step description of each algorithm, we have included programs which factor numbers up to 19 digits using the Hewlett-Packard HP-16C. This calculator has a 64-bit word size and built-in double-precision multiplication and remainder (or quotient) on division. These programs and instructions for use are formatted so that they can be photocopied, mounted or laminated, and carried in the calculator case for easy access. The algorithms given here would only work up to about five or six digit numbers on other programmable calculators.

The strategy. After dividing out any power of two, we may assume that the number N which we wish to factor is odd. We start with "baby divide" which simply divides N by successive odd numbers and halts when it finds a factor. This program is time-consuming and should only be used to take out small factors. Selfridge and Guy **[5]** recommend using baby divide to find factors up to about ten times the number of decimal digits of N.

After removing any small factors, we use the power algorithm described below to compute $2^{N-1} \pmod N$. Fermat's little theorem asserts that $2^{N-1} \equiv 1 \pmod N$ for any odd prime N. If this congruence does not hold, then N is composite, and we apply the Pollard rho algorithm to find two factors of N which may or may not be prime. However, we do not have to check primality for any factor which is smaller than the square of the largest divisor tried in baby divide, since such a factor is necessarily prime.

If, on the other hand, $2^{N-1} \equiv 1 \pmod N$, then N is probably prime, and to confirm this we apply the Lucas test. Define a Lucas sequence by $U_0 = 0$, $U_1 = 1$, $U_{n+1} = U_n - QU_{n-1}$ for fixed Q (not 0 or 1). The following theorem is an analogue of Fermat's little theorem: *If N is prime, $N > Q$, and the Jacobi symbol $((1 - 4Q)/N) = -1$, then $N \mid U_{N+1}$.* The Lucas test for the number N then consists of first finding small integers D and Q such that $(D/N) = -1$, where $D = 1 - 4Q$, and then checking to see if $U_{N+1} \equiv 0 \pmod N$. If this congruence does not hold, then N is composite* (go to Pollard rho), but if it does hold then N is almost certainly prime. In fact, if the D is chosen as suggested in our discussion of the algorithm, and N passes both the Fermat test $2^{N-1} \equiv 1 \pmod N$ and the Lucas test $U_{N+1} \equiv 0 \pmod N$, then Pomerance, Selfridge and Wagstaff **[4]** have shown that N is prime for any $N < 25 \cdot 10^9$. Even if $N > 25 \cdot 10^9$, there is

*Composite numbers for which $N \mid 2^N - 2$ are called pseudoprimes (base 2). They are much rarer than primes.

no known composite N which passes the two tests. In their paper, Pomerance, Selfridge and Wagstaff offer \$30, since increased to \$120, for the first submission of such a composite N or for a proof that none exists.

The algorithms

Baby divide.

Input N	4. If $f \nmid N$, go to 3
1. $3 \to f$	5. $N/f \to N$
2. Go to 4	6. Halt showing f
3. $f + 2 \to f$	7. Go to 4

Power algorithm: $a^E \pmod N$. In order to compute 3^{22} we could perform 21 multiplications by 3, but a faster approach is to compute 3^1, 3^2, 3^5, 3^{11}, 3^{22}, the exponents in binary being 1, 10, 101, 1011 and 10110. Each step is a squaring, and we also multiply by 3 when the new binary digit is a one.

In general, to compute a^E we express E in binary. Then, examining E left to right and starting with $R = 1$, we square the current value of R and multiply by a when we encounter a one bit and merely square R when we encounter a zero bit. Usually E will end with one or more zero bits, and it is convenient in our algorithm to annex a signal bit 1 at the right of E. We shift E left one place at each iteration and simply check for zero to see when we have finished.

Power algorithm: $a^E \pmod N$.

Input $N, a, E = (b_m b_{m-1} \cdots b_0)_2$	6. If $C = 0$, go to 8
1. $2E + 1 \to E$ (annex trailing bit 1)	7. $aR \pmod N \to R$ (shift left when $a = 2$)
2. Shift E left until bit shifted out is 1	8. Shift E left one place
	9. Bit shifted out $\to C$
3. $1 \to C$	10. If $E \neq 0$, go to 5
4. $1 \to R$	11. Halt showing R
5. $R^2 \pmod N \to R$	

Pollard rho. The Pollard rho method [3] gets its name from Pollard and from the fact that if

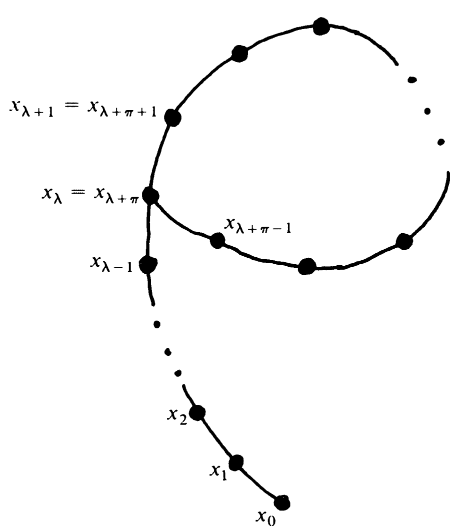


FIG. 1

we iterate a function f from a finite set into itself, $x_{n+1} = f(x_n)$, then there exist positive integers λ and π such that $x_{\lambda+j} = x_{\lambda+\pi+j}$ for all nonnegative integers j. The least such λ and π are called the tail length and the period, respectively, of the function. The picture we get is given in Fig. 1 and resembles the letter rho.

We apply this simple observation to factor N by noting that if the prime p divides N, and if we recursively apply $f(x) = x^2 + a$ to the integers modulo p starting with x_0, then eventually $x_h \equiv x_k \pmod p$ and so $p \mid \mathrm{GCD}(x_h - x_k, N)$. Of course we do not know p, but we have noted that $\mathrm{GCD}(x_h - x_k, N)$ is a factor of N greater than 1 for some h and k. Next we note that if $x_h \equiv x_k \pmod N$ then $x_h \equiv x_k \pmod p$, and so we can keep track of $x_h - x_k \pmod N$ even though we don't know p. We observe that if $x_h \equiv x_k \pmod p$, then $x_{h+j} \equiv x_{k+j} \pmod p$ for all positive integers j. Thus, we are only interested in the difference of the indices $h - k$. Since it is not necessary to find λ and π, we simply compute $\mathrm{GCD}(x_h - x_k, N)$ for $h - k = m + 1, m + 2, \ldots$. We expect the prime p to appear within about $p^{1/2}$ iterations. It can be seen from the table at the end of the paper that any prime shows up in a reasonable time, using our chosen function.

We begin computing $x_h - x_k$ with $k = 0$ and $h = m + j$ for $j = 1, 2, \ldots, m$. Then we let $k = 2m$ and $h = 4m + j$ for $j = 1, 2, \ldots, 2m$, and so on. For the tth iteration $k = (2^t - 2)m$ and $h = k + 2^{t-1}m + j$ for $j = 1, 2, \ldots, 2^{t-1}m$. In this way, we have $h - k$ take every value from $m + 1$ onward, and at each iteration we advance the smaller index forward along the tail toward the periodic part of the rho. Thus we eventually have both indices larger than λ, and even if we do not have $k \geqslant \lambda$ when $h - k = \pi$, we will have $k \geqslant \lambda$ for $h - k$ equal to some multiple of π. To speed things up, we do not compute the GCD for each $x_h - x_k$, but rather we form the product $\pmod N$ of m consecutive $x_h - x_k$ and then compute the GCD. (This makes it convenient to start $h - k$ at $m + 1$.) We have used $m = 8$ in our HP-16C program.

The possibility exists that when we find a GCD greater than 1, it may turn out to be N. Fortunately, this happens only rarely and almost never after a long computation. Although an obvious strategy would be to go back and repeat the last cycle of m differences, computing the GCD for individual $x_h - x_k$ rather than for the product, we do not do this since we are too short of space in the HP-16C. Even if we did this individual check, we might still have the GCD equal to 1 or N for each $x_h - x_k$. We suggest that the a in $f(x) = x^2 + a$ be increased by 1, and the Pollard rho run again.

The version of the Pollard rho algorithm that we have used follows modifications due to Brent [1]. Originally Pollard used $x_{2k} - x_k \pmod N$, but this meant that both terms had to be advanced when we go to $x_{2k+2} - x_{k+1}$. Brent's modification was found to be about 24% faster than the original.

Pollard rho.

Input N, a, m
1. $X_0 \to X$
2. $m/2 \to I$
3. $2I \to S, J, I$
4. $X \to Y$
5. $X^2 + a \pmod N \to X$
6. $S - 1 \to S$
7. If $S \neq 0$, go to 5
8. If $J = 0$, go to 3
9. $m \to S$
10. $J - m \to J$
11. $1 \to R$
12. $X^2 + a \pmod N \to X$
13. $|X - Y| \cdot R \pmod N \to R$
14. $S - 1 \to S$
15. If $S \neq 0$, go to 12
16. $\mathrm{GCD}(N, R) \to D$
17. If $D = 1$, go to 8
18. Halt showing D
19. Show N/D

GCD(N, R) ($N > R \geqslant 0$).

1. $R \to X$
2. $N - R \to Y$
3. $X \rightleftarrows Y$ (swap X and Y)
4. $Y \pmod X \to Y$
5. If $Y \neq 0$, go to 3
6. Return showing X

Lucas test.

To apply the Lucas test, we pick an appropriate Q (and D) and compute U_{N+1} (mod N). To get Q, we first find the least D in the sequence 5, -7, (9), -11, 13, $-15, \ldots$ such that $(D/N) = -1$ by using the elementary properties of the Jacobi symbol. Then $Q = (1 - D)/4$. We have included in the program description a table of Q which works for 99.2% of N's.

To compute U_{N+1} we define the auxiliary sequence $V_0 = 2$, $V_t = U_{2t}/U_t$. The following formulas are well known:

Doubling Formulas: $U_{2t} = U_t V_t$ and $V_{2t} = V_t^2 - 2Q^t$.

Sidestep Formulas: $U_{2t+1} = (U_{2t} + V_{2t})/2$ and $V_{2t+1} = (DU_{2t} + V_{2t})/2$.

Starting with $U_0 = 0$ and $V_0 = 2$, the sequence of doublings and sidesteps necessary to compute U_{N+1} and V_{N+1} is obtained from the binary expansion of $N + 1$, just as we handle E in the algorithm for a^E(mod N).

Lucas test.

Input $N, Q, I = N + 1 = (b_m b_{m-1} \cdots b_0)_2$
1. $1 \to R$
2. $2 \to V$
3. $0 \to U$
4. $2I + 1 \to I$ (annex trailing bit 1)
5. Shift I left until leftmost bit is 1
6. UV (mod N) $\to U$
7. V^2 (mod N) $- 2R \to V$
8. R^2 (mod N) $\to R$
9. Shift I left one place
10. Bit shifted out $\to C$
11. If $C = 0$, go to 16
12. $U \to T$
13. $(U + V)/2$ (mod N) $\to U$
14. $((1 - 4Q)T$ (mod N) $+ V)/2$ (mod N) $\to V$
15. QR (mod N) $\to R$
16. $I \to X$
17. Shift X left one place
18. If $X \neq 0$, go to 6
19. Halt showing U

Factoring programs for the HP-16C

We include the HP-16C code for implementing the above algorithms. The main reason for presenting the actual code is that one must be careful when writing these programs to take full advantage of the HP-16C's 19-digit capacity. After the necessary 40 bytes are set aside for storing five 19-digit numbers, there are 161 bytes remaining for program storage. The programs presented here use 159 of these bytes. (See Fig. 2 on p. 806.)

Using unsigned mode we can handle numbers up to 2^{64} in baby divide and Pollard rho. In our program for a^E (mod N), when $N > 2^{62}$ we must store a in the I register and have GSB F in 025; also E must be less than 2^{63}. This forces us to check $a^{(N-1)/2} \equiv 1$ or $N - 1$ (mod N) in the Fermat test when $N > 2^{63}$. In the Lucas test we must be in 2's complement mode with $N < 2^{63}/5$.

In these programs labels 9 and A are not used, and label F is used several times "locally". By changing the word size, storage can be made available for short temporary programs without disturbing the factoring package.

Two other programs. We have also written programs for division or multiplication of a number having up to 396 digits by factors up to 2^{64}. These two programs can be obtained by writing to us.

EXAMPLE 1. We enter 1542 74344626 34653133 into the machine. (Since $N > 2^{63}$, we use unsigned integer mode.) First we use baby divide which finds the factor 17 in 12 sec., a second 17 in two more sec., and the factor 101 a minute later. Since the remaining cofactor is a 15-digit number, we continue baby divide for 40 sec. longer, trying all odd divisors less than 159. Next we

B Baby divide

Use Baby divide to find small odd factors. Go up to about ten times the number of decimal digits of N.

Storage: N in X; STO 0; f_0 (= 3) in X. GSB B.

Halts showing factor. Record and GSB B to continue (N has been replaced by N/f).

To show current f: R/S, LSTX.
To continue: R↓, R/S.

C Pollard rho $x^2 + a$

If $2^{N-1} \not\equiv 1$ (mod N) or $U_{N+1} \not\equiv 0$ (mod N), use Pollard rho to find factors of N. First check MEM. If r < 5, see directions in Lucas test.

Storage: N in 0; 2 in 2 (x_0); 4 in 4 (2^t); a (= 1) in program step 082. GSB C. Halts with factors in X and Y. If result is N and 1, increase a and try again.

Note: Subroutine D may be used alone to find GCD's. Compute the difference (in the X register) of the two numbers for which you wish to find the GCD and GSB D.

E a^E mod N

After using Baby divide to take out small factors, use a^E mod N with $a = 2$ (SL in program step 025) and E = N−1 to determine whether N is a probable prime.

Storage: N in 0; E (= N − 1) in X. GSB E. (For $a > 2$ or E odd, put a in register I and change step 025 from SL to GSB F. When step 025 contains SL, E should be even.)

Halts showing result. If $2^{N-1} \equiv 1$ (mod N), N is probably prime. (See Lucas test.)

2 Lucas test: **change to 2's complement mode**

If $2^{N-1} \equiv 1$ (mod N), N is a probable prime; use Lucas test to verify primality. First check MEM. If r < 5, delete temporary programs or use smaller word size so that r ≥ 5.

Storage: N in 0; 1 in 1 (Q^k); 2 in 2 (V_k); 0 in 3 (U_k); Q (= (1 − D)/4) in 4; final subscript (N + 1) in X. GSB 2.

Halts showing U_k (mod N). (k = N + 1)

If $U_{N+1} \not\equiv 0$ (mod N), N is composite. Go to Pollard rho. If $2^{N-1} \equiv 1$ (mod N) and $U_{N+1} \equiv 0$ (mod N), N is prime if $N < 25 \cdot 10^9$.

B Baby divide 001–012

LBL 1
RCL 0
2
LSTX
+
LBL B
÷
F? 4
GTO 1
STO 0
LSTX
RTN

E a^E mod N 013–034 (and 085–091)

LBL E	
SF 4	SL (GSB F)
RLC	R↑
LJ	SL
R↓	X≠0
SL	GTO 7
1	X≷Y
X≷Y	RTN
LBL 7	
X≷Y	LBL F
GSB 3	RCL I
F? 4	GTO 4

C Pollard rho 035–099

LBL C	STO 2	GSB F	GSB D	1 (= a)	LBL D
RCL 4	LBL 0	STO 2	1	+	LSTX
SL	RCL 3	RCL 1	X=Y	RTN	X≷Y
STO I	X=0	X>Y	GTO 0		RMD
STO 3	GTO C	X≷Y	RCL 0	LBL 3	X≠0
STO 4	8	−	LSTX	ENTER	GTO D
RCL 2	STO I	GSB 4	÷	LBL 4	LSTX
STO 1	−	DSZ	LSTX	DBLx	RTN
LBL 6	STO 3	GTO 8	RTN	RCL 0	
GSB F	1	RCL 0		DBLR	
DSZ	LBL 8	X≷Y	LBL F	RTN	
GTO 6	RCL 2	−	GSB 3		

2 Lucas test ($N < 2^{63}/5$): Q Tables

If N ends in 3 or 7, Q = −1.
If N mod 7 is 3, 5, or 6, Q = 2.
If N mod 11 is 2, 6, 7, 8, or 10, Q = 3.
If N mod 13 is 2, 5, 6, 7, 8, or 11, Q = −3.
If N mod 15 is 7, 11, 13, or 14, Q = 4.
If N mod 17 is 3, 5, 6, 7, 10, 11, 12, or 14, Q = −4.
If N mod 19 is 2, 3, 8, 10, 12, 13, 14, 15, or 18, Q = 5.

If Q is not found using any of these moduli, use the following: D is the first non-square element of the sequence 5, −7, (9), −11, 13, −15, ⋯ for which (D/N) = −1. Then Q = (1 − D)/4.

2 Lucas test 100–159 (and 085–091)

LBL 2	GSB 3	GSB F	LBL F	−	LBL F
SF 4	RCL 1	RCL I	RCL 3	GSB 4	RCL 2
RLC	SL	SL	RCL 0	GSB F	+
LJ	−	X≠0	RCL 3	STO 2	0
R↓	STO 2	GTO 5	GSB F	RCL 1	B?
STO I	RCL 1	RCL 3	STO 3	RCL 4	+
LBL 5	GSB 3	RTN	RCL 0	GSB 4	ASR
RCL 3	STO 1		R↑	STO 1	RTN
RCL 2	RCL I		1	RTN	
GSB 4	SL		RCL 4		
STO 3	STO I		SL		
RCL 2	F? 4		SL		

FIG. 2

use a^E mod N with $a = 2$. We find $2^{N-1} \not\equiv 1 \pmod N$ in two min. Thus the remaining number is composite, so we use Pollard rho to find the factors 23209 and $N = 227\ 72885633$ in 29 min. Since $23209 < 159^2$, we know that 23209 is prime. We next test N using a^E mod N. We find $2^{N-1} \equiv 1 \pmod N$ in 1.5 min., and then proceed to the Lucas test. Because N ends in 3, $Q = -1$. The Lucas test takes 5.5 min. and shows $U_{N+1} \equiv 0 \pmod N$. Since $N < 25 \cdot 10^9$, we are sure that it is prime. Thus

$$1542\ 74344626\ 34653133 = 17^2 \cdot 101 \cdot 23209 \cdot 22772885633.$$

The complete factorization is accomplished in about 40 min.

Advanced Pollard rho. When applying the Pollard rho algorithm to a composite N, it is not necessarily the case that the factor D of N, which is found first, is the smallest factor of N, and indeed it may not even be prime. If D is not prime, its prime factors will not show up using $x^2 + a$ with the current value of a. However, this value of a is probably still good for finding further factors of $M = N/D$. If M is composite, we should continue the Pollard rho algorithm on M with the parameters at those values where D was found. Thus, when D appears, we set it aside for further work later, and first do a Fermat test on M. If $2^{M-1} \equiv 1 \pmod M$, we confirm the primality of M by a Lucas test. If $2^{M-1} \not\equiv 1 \pmod M$, we reduce x_k and x_h (mod M) and continue Pollard rho working on M with these values of x_k and x_h. Later when we return to consider the factor D we do a Fermat test and, if need be, a Lucas test. If D is not prime, we have a choice: continue baby divide until it finds a factor or increase the current value of a and start Pollard rho from the beginning.

EXAMPLE 2. Consider $N = 750\ 05962469\ 54111183$. After running baby divide for 2.5 minutes, we have tried all potential odd factors up to 200 and found none. After 3.25 minutes on a Fermat test, we know that N is not prime. (Note that when we use $a^E \pmod N$ on an $N > 2^{62}$ we must have GSB F in 025 and a in the I register. We then check whether $a^{(N-1)/2} \equiv 1$ or $N - 1 \pmod N$. We remember to restore SL in 025 when this task is done.) The Pollard rho algorithm finds the factor $D = 3350797$ after 7.25 minutes. (It is surprising to see such a large factor in so short a time.) We write down D for consideration later and apply a Fermat test on $M = N/D$ by simply executing R$\downarrow$, STO 0, 1, $-$, GSB E. In 1.7 minutes, we observe that $2^{M-1} \not\equiv 1 \pmod M$, and so we reduce the current x_k and x_h modulo M by executing RCL 1, RCL 0, RMD, STO 1; RCL 2, RCL 0, RMD, STO 2, and then continue Pollard rho by GSB 0. After 1.7 minutes, the factor 24977 appears, and it is necessarily prime since it is less than 200^2. We determine that the cofactor 89620507 is prime by the Fermat and Lucas tests in 5 minutes. We next return to 3350797 and run a Fermat test, determining that it is composite (1 min.). We then change $a = 1$ in 082 to $a = 2$ and run Pollard rho, finding the prime factors 1873 and 1789 in 5 minutes. Thus

$$750\ 05962469\ 54111183 = 1789 \cdot 1873 \cdot 24977 \cdot 89620507.$$

The complete factorization is accomplished in less than half an hour. However, for some stubborn large numbers, you have to let Pollard rho run overnight (in the worst possible case even longer). The machine turns off the power soon after finding the factors, and you have them in the morning.

Just as it is unnecessary to check the primality of any factor found which is less than the square of the largest divisor used in baby divide, it is often unnecessary to check the primality of one (or even both) Pollard rho factors. Specifically, when Pollard rho halts note the power of two, 2^t, in register 4. If we let p_t denote the least prime for which register 4 is equal to (or greater than) 2^t when the prime is found, then N has no prime factor smaller than p_t. Hence any factor D of N found with 2^t in register 4 must be prime if $D < p_t^2$. We have included a table of p_t for the function $x^2 + 1$.

Primality table for $a = 1$

2^t	p_t	$p_t^2 - 1$
32	193	37248
64	607	368448
128	1747	3052008
256	11261	1 26810120
512	21911	4 80091920
1024	100417	100 83573888

Any factor less than p_t^2 is prime.

References

1. R. P. Brent, An improved Monte Carlo factorization algorithm, BIT, 20(1980) 176–184.
2. John Brillhart, D. H. Lehmer, and J. L. Selfridge, New primality criteria and factorizations of $2^m \pm 1$, Math. Comp., 29(1975) 620–647.
3. J. M. Pollard, A Monte-Carlo method for factorization, BIT, 15(1975) 331–334.
4. Carl Pomerance, J. L. Selfridge, and Samuel S. Wagstaff, Jr., The Pseudoprimes to $25 \cdot 10^9$, Math. Comp., 35(1980) 1003–1026. (The $30 prize on page 1025 has been upped to $120.)
5. J. L. Selfridge and Richard K. Guy, Primality Testing with Application to Small Machines, Proc. WSU Conf. Number Theory, Pullman, WA, 1971, 45–51.

THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

For instructions about submitting material for publication in this department see the inside front cover.

A "GREAT THEOREMS" COURSE IN MATHEMATICS

WILLIAM DUNHAM
Department of Mathematics, Hanover College, Hanover, IN 47243

I recently came across an advertisement for a new textbook on number theory. The ad's main thrust was to acclaim the applications of number theory in the construction of unbreakable codes. In essence, it implied that such a real-world application had at last legitimized number theory as a respectable area of human inquiry.

As I tossed the advertisement aside, I was left feeling decidedly uneasy. Even number theory —long regarded as the purest of the pure—must now be marketed as applied mathematics, somehow vital to the national security, in order to be considered worthwhile. Of course, it was not its strategic utility that led Euclid or Fermat or Euler or Gauss to devote so much of their energy and genius to the "higher arithmetic." These gentlemen did not feel compelled to justify their number theoretic work by its real-world applications, any more than Shakespeare had to apologize for writing love sonnets instead of cookbooks or than Van Gogh had to apologize for painting canvases instead of billboards.

Of course, applications are inseparable, both logically and historically, from mathematics itself, and recent advances in applied number theory are surely interesting. Yet I am afraid that the mathematics professor of today tends to ignore the centuries-old conception of mathematics as an artistic exercise in pure reason that can provide fascinating glimpses into the logical relationships of number and space. Too often, we fall into the easy trap of stressing only the applications of mathematics to commerce or computers. In so doing, I think we ignore an

CENTER SECTION

(Vol. 93, No. 10, December 1986)

Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T:	Textbook	P:	Professional Reading	1-4:	Semesters
C:	Computer Software	L:	Undergraduate Library	** :	Special Emphasis
S:	Supplementary Reading	13:	Grade Level	?? :	Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the Monthly.

General, S*, L*. Mathematical Quickies: 270 Stimulating Problems with Solutions. Charles W. Trigg. Dover, 1985, xi + 210 pp, $4.95 (P). [ISBN: 0-486-24949-2] An unabridged and corrected reproduction of the 1967 work first published by McGraw-Hill (TR, March 1968). LCL

General, P, L*. Jean-Pierre Serre: Oeuvres, Collected Papers. Jean-Pierre Serre. Springer-Verlag, 1986, $198 set [ISBN: 0-387-15621-6]. Volume I: 1949-1959, xviii + 596 pp; Volume II: 1960-1971, iv + 740 pp; Volume III: 1972-1984, iv + 728 pp. 132 papers arranged chronologically, including almost all of Serre's published papers, a selection of his seminar notes, and summaries of his courses at Collège de France since 1956. Each volume includes a complete list of Serre's works including books, seminar papers and several joint papers not included here. LAS

Precalculus, T(13: 1). Algebra and Trigonometry. Howard A. Silver. Prentice-Hall, 1986, xv + 654 pp, $29.95. [ISBN: 0-13-021270-9-01] Traditional precalculus including trigonometry and algebra with good examples accompanied by flow chart descriptions of the process, noting common errors. Easy-to-read layout and many problems. Built-in use of calculator to simplify and reinforce concepts. GF

Precalculus, T(1, 2). Algebra and Trigonometry. David Cohen. West, 1986, xi + 657 pp, $31.95. [ISBN: 0-314-93165-1] In addition to the expected material, includes some elementary matrix theory, theory of equations, and special topics (mathematical induction, binomial theorem, sequences and series, permutations and combinations, probability). Thousands and thousands of exercises, many designed for solution using calculators. Assumes only high school intermediate algebra. DFA

Precalculus, T(13: 1). Essentials of Trigonometry, Second Edition. Karl J. Smith. Brooks/Cole, 1986, xi + 292 pp, $29.50. [ISBN: 0-534-05274-6] Appears to be a reprint of the First Edition (TR, October 1983). LCL

Precalculus, T(13: 1, 2). Basic Technical Mathematics. Charles R. Wall. Harcourt Brace Jovanovich, 1986, xii + 745 pp, $32.95. [ISBN: 0-15-505130-X] Covers main precalculus topics of algebra, trigonometry, and analytic geometry, plus material on descriptive statistics and an introduction to limits. Close to 4,000 exercises (many set in engineering situations); more than usual emphasis on applications; stress on calculator usage. LCL

Logic, P. Lecture Notes in Mathematics-1141: Recursion Theory Week. Ed: H.-D. Ebbinghaus, G.H. Müller, G.E. Sacks. Springer-Verlag, 1985, ix + 418 pp, $25.80 (P). [ISBN: 0-387-15673-9] Proceedings of a conference held in Oberwolfach, West Germany, April 15-21, 1984. Twenty research papers plus article by Sacks presenting 49 open questions. KS

Logic, P. Theory of Relations. R. Fraïssé. Stud. in Logic & Found. of Math., V. 118. Elsevier Science, 1986, xii + 397 pp, $55.25. [ISBN: 0-444-87865-3] Relation theory originates in the theory of order types, but the subject matter intersects with graph theory, mathematical logic, set theory, combinatorics, and topology. The first eight chapters present the principal notions and their present state, and the final four chapters concern the general theory. LCL

Foundations, P. The Prolegomena to a 1985 Philosophiae Naturalis Principia Mathematica. Filmer Stuart Cuckow Northrop. Ox Bow Pr, 1985, xvi + 73 pp, $29.95. [ISBN: 0-918024-35-8] A preface to a proposed four-volume presentation of the "science of the true" on mathematical physics, aesthetics, jurisprudence and theology. Obscurely written, this preface discusses philosophical links concerning "first principles" among Euclid, Newton, Darwin, and especially Einstein, relating these ideas by vague allusions to a variety of philosophers, ancient and modern. Claims to be the first really new theory of knowledge since Kant. LAS

Graph Theory, T*(15: 1, 2), S, P, L*. Graphs & Digraphs, Second Edition. Gary Chartrand, Linda Lesniak. Math. Ser. Wadsworth, 1986, viii + 359 pp, $39.15. [ISBN: 0-534-06324-1] A major change in this edition is the integration of graph and digraph theory. In addition, this edition includes problems that can be modelled by graphs together with efficient algorithms for their solutions and careful proofs that the algorithms work. Good problem sets and list of references. (First Edition, TR, October 1979.) CEC

Combinatorics, P. Algorithms in Combinatorial Design Theory. Ed: E.J. and M.J. Colbourn. Math. Stud., V. 114. Elsevier Science, 1985, viii + 334 pp, $45 (P). [ISBN: 0-444-87802-5] Collection of 16 papers elucidating algorithmic aspects of combinatorial design theory which include generation, isomorphism, analysis techniques, and computational complexity of these operations. SS

Discrete Mathematics, T(13-15: 1, 2). Applied Discrete Structures for Computer Science. Alan Doerr, Kenneth Levasseur. SRA, 1985, xx + 523 pp, $32.95. [ISBN: 0-574-21755-X] Set theory and logic, relations and function, recursion and recurrence relations, graph theory and trees, matrix algebra, algebraic structures (Boolean algebra, monoids, groups, rings and fields). Unique features: wide coverage, major theoretical topics are reinforced with applications to computer science, "Pascal Notes" (optional discussions relating to Pascal and other programming languages). LCL

Number Theory, P. Thirteen Papers in Algebra and Number Theory. I.K. Zhuk, et al. AMS Transl., Ser. 2, V. 128. AMS, 1986, v + 122 pp, $49. [ISBN: 0-8218-3097-X] An assortment of papers on subjects in algebraic number theory and related algebraic problems. BC

Number Theory, S*(15-17), P, L.** Number Theory in Science and Communication: With Applications in Cryptography, Physics, Digital Information, Computing, and Self-Similarity, Second Enlarged Edition. M.R. Schroeder. Ser. in Inform. Sci., V. 7. Springer-Verlag, 1986, xix + 374 pp, $24.50 (P). [ISBN: 0-387-12164-1] Additions to the 1984 First Edition (TR, January 1985) include quasicrystals, self-similarity, fractals, chaos, and error-free computation. A marvelous exploration of uses of number theory in a broad range of mathematical, physical and communication sciences, supposedly written for non-mathematicians but in reality requiring solid undergraduate mathematical background. A delightful and complete refutation of Hardy. LAS

Linear Algebra, T*(14: 1), S, L. Elementary Linear Algebra, Fourth Edition. Bernard Kolman. Macmillan, 1986, xv + 389 pp. [ISBN: 0-02-366080-5] This edition has an appendix which introduces complex numbers in linear algebra, additional material on projections, more exercises, illustrative examples, more figures, unifying summaries of important results and rewriting for the sake of clarity in several places. The notation det(A) has also been introduced. (First Edition, TR, June-July 1970 and January 1971; Extended Review, March 1974; Second Edition, TR, April 1977; Third Edition, TR, January 1983.) CEC

Group Theory, P, L*.** Atlas of Finite Groups: Maximal Subgroups and Ordinary Characters for Simple Groups. J.H. Conway, et al. Clarendon Pr, 1985, xxxiii + 252 pp. [ISBN: 0-19-853199-0] Intended to convey "every interesting fact about every interesting finite group" by providing computer-generated tables for finite simple groups containing various constructions, information about subgroups and automorphism groups, and a "compound character table." This information is displayed on oversized 12" by 16" pages, covering all 26 sporadic groups and the early terms in the infinite families "until they become either too big or too boring." Begins with a concise overview of simple groups and instructions for reading (decoding) the idiosyncratic data on the nearly 100 group displays. LAS

Topological Groups, T(17-18: 1, 2), S, P. Lie Groups and Algebras with Applications to Physics, Geometry, and Mechanics. D.H. Sattinger, O.L. Weaver. Appl. Math. Sci., V. 61. Springer-Verlag, 1986, ix + 215 pp, $29.80. [ISBN: 0-387-96240-9] Written for research mathematicians and physicists, but suitable as an introductory text. (Good examples and numerous exercises.) Basic topics: Lie groups and algebras, differential geometry and Lie groups, algebraic theory, and representation theory. The "language" of differential forms is stressed. BC

Topological Groups, P. Analysis on Non-Riemannian Symmetric Spaces. Mogens Flensted-Jensen. CBMS Reg. Conf. Ser. in Math., No. 61. AMS, 1986, x + 77 pp, $13 (P). [ISBN: 0-8218-0711-0] Well-written treatment of pseudo-Riemannian symmetric spaces with semi-simple isometry group. Requires familiarity with semi-simple Lie groups and algebras. Based on expository lectures given at a CBMS conference in 1984. BC

Topological Groups, T(18: 1, 2), S, P. Infinite Dimensional Lie Algebras, Second Edition. Victor G. Kac. Cambridge U Pr, 1985, xvii + 280 pp, $24.95. [ISBN: 0-521-32133-6] A clear account of Kac-Moody algebras by one of the founders. The introduction alone is worth reading. Eminently suitable as an introduction (assuming familiarity with semi-simple Lie algebras), with a surprising number of exercises. (First Edition, TR, January 1985.) BC

Algebra, P. Radical Theory. Ed: L. Márki, R. Wiegandt. Elsevier Science, 1985, 753 pp, $74. [ISBN: 0-444-86765-1] The proceedings (surveys and refereed research papers with full proofs) of a conference on radical theory held in Eger, Hungary, August 1-7, 1982, attended by 64 algebraists from 22 countries. Includes a collection of open problems presented by the participants at the problem session. LCL

Calculus, T(13: 2). Calculus for Business, Economics, and the Social and Life Sciences, Third Edition. Laurence D. Hoffmann. McGraw-Hill, 1986, xii + 708 pp, $32.95. [ISBN: 0-07-029331-7] A cal-

culus text which is application-oriented for those going into social, managerial, and life sciences, especially business and economics. New edition includes expanded coverage of trigonometry and inclusion of topics on improper integrals, double integrals, infinite series, numerical integration methods, Taylor's approximation and Newton's method and method of least squares, and more economic applications. Intuitive and geometric understanding of concepts with problems in application to many practical situations. GF

Calculus, T(13: 2), S. An Introduction to Calculus: Methods and Applications. James R. Evans, Charles W. Groetsch, Maryanne Walker. West, 1986, xix + 658 pp, $34.95. [ISBN: 0-314-93176-7] A brief, intuitive introduction to methods and applications of calculus and elementary probability. Little theory; e.g., mean value theorems are omitted, fundamental theorem not proved. Emphasizes word problems (especially in economics), graphical, tabular, and algebraic functions. Few scientific applications. Chapter introductions pose problems requiring techniques to follow. Includes several BASIC programs for numerical methods. PZ

Real Analysis, T*(14: 1), S, L*. Convergence, Approximation, and Differential Equations. Eugene A. Herman. Wiley, 1986, xv + 380 pp, $34.95. [ISBN: 0-471-81762-7] An innovative approach for integrating discrete and continuous concepts into the sophomore curriculum. The course offers the instructor considerable flexibility for mixing theory, computation, and application; the book is written so that each of these aspects support the others. Written for a fourth-term course (preceded by two terms of calculus, and one term of linear algebra and differential equations), the topics include sequences, numerical approximation, series, series solutions to differential equations, Fourier series. LCL

Real Analysis, T*(16-17: 2), L*. Probability and Measure, Second Edition. Patrick Billingsley. Appl. Prob. & Math. Stat. Wiley, 1986, xii + 622 pp, $44.95. [ISBN: 0-471-80478-9] A revision of the successful and popular 1979 edition (TR, November 1981), expanded to include, among other things, a central limit theorem for Martingales and more on the asymptotic behavior of Markov chains. In this year-long course the student is exposed to the fundamental ideas of both subjects in about equal doses, and should, after working many of the 600+ exercises, be well prepared to do advanced work in either. TAV

Real Analysis, T*(16-17: 1), S, L. Measure Theory and Probability. Malcolm Adams, Victor Guillemin. Wadsworth, 1986, xii + 203 pp, $32.40. [ISBN: 0-534-06330-6] Intended as a one-semester course in measure theory for good undergraduates. The authors use probability as a framework to study measure and integration. With the probability models in mind, "the students [are] much better able to endure the long arid trek through the basics of measure theory." In fact, the text is neither long nor arid, but lively and concise with numerous exercises and examples, providing an interesting and understandable first view of the basics of measure and integration. TAV

Complex Analysis, T(16-17: 1), S, L.** Complex Variables. Stephen D. Fisher. Math. Ser. Wadsworth, 1986, xii + 403 pp, $39.90. [ISBN: 0-534-06168-0] Introduction to theory, methods, and mathematical applications of analytic and harmonic functions. Exposition is readable and friendly, yet mathematically elegant and enthusiastic. Problem sets alone--excellent, varied, many with annotated exercises introducing related topics--are worth the price. Stresses analogy to real calculus, e.g., in treating series, line integrals. Many references, figures. PZ

Complex Analysis, S(18), P. Two Papers on Extremal Problems in Complex Analysis. S. Ya. Khavinson. AMS Transl., Ser. 2, V. 129. AMS, 1986, v + 114 pp, $50. [ISBN: 0-8218-3099-6] Two long expository papers on function-theoretic foundations of extremal problems for bounded analytic functions of one variable. The author's duality method is key to many results. Russian originals published in 1981. PZ

Complex Analysis, P. Analysis on Real and Complex Manifolds. R. Narasimhan. Math. Lib., V. 35. Elsevier Science, 1985, xiv + 246 pp, $55. [ISBN: 0-444-87776-2] Third printing, with a new preface, of the 1968 original edition (TR, October 1969), based on 1964-65 Tata Institute lectures. First chapter covers main theorems of differential topology; second chapter introduces real and complex manifolds; third chapter studies elliptic differential operators. PZ

Partial Differential Equations, P. Nonlinear Systems of Partial Differential Equations in Applied Mathematics. Ed: Basil Nicolaenko, Darryl D. Holm, James M. Hyman. Lect. in Appl. Math., V. 23. AMS, 1986, $80 set. Part 1, xvi + 468 pp [ISBN: 0-8218-1125-8]; Part 2, x + 387 pp. [ISBN: 0-8218-1126-6] Collection of papers prepared for the SIAM-AMS summer seminar on systems of non-linear partial differential equations. Papers are grouped generally into five sections: integrable systems, variational problems, dispersive systems, evolutionary systems, and hyperbolic systems. AM

Partial Differential Equations, P. Recent Topics in Nonlinear PDE. Ed: Masayasu Mimura, Takaaki Nishida. Math. Stud., V. 98. Elsevier Science, 1984, vii + 239 pp, $46.25 (P). [ISBN: 0-444-87544-1] Collection of papers given at a meeting on nonlinear partial differential equations held at Hiroshima University in February 1983. Topics covered include fluid dynamics, free boundary problems, population dynamics, and mathematical physics. AM

Partial Differential Equations, T*(15-17: 1, 2). Elementary Partial Differential Equations with Boundary Value Problems. Larry C. Andrews. Academic Pr, 1986, xii + 520 pp, $27. [ISBN: 0-12-059510-9] Excellent introduction to partial differential equations and boundary value problems. Would be an appropriate text to follow an undergraduate course in ordinary differential equations taught at the level of Boyce and DiPrima. Book presents standard solution techniques using both

theory and physical examples. Introduces delta functions, but makes no systematic use of distributions. Contains many exercises with answers for odd numbered problems. AM

Numerical Analysis, P. Computer Methods for the Range of Functions. H. Ratschek, J. Rokne. Ser. in Math. & Its Applic. Halsted Pr, 1984, 168 pp, $42.95. [ISBN: 0-470-20034-0] Concentrates on the use of interval arithmetic and various centered forms for approximating ranges. RWN

Numerical Analysis, P. Multigrid Methods for Integral and Differential Equations. Ed: D.J. Paddon, H. Holstein. Inst. of Math. & Its Applic., Conf. Ser, V. 3. Oxford U Pr, 1985, xii + 323 pp, $47.50. [ISBN: 0-19-853606-2] 12 papers from a conference held in Bristol, U.K. Includes spectral, hierarchical finite element and algebraic multi-grid methods and applications to engineering design, simulation and elliptic problems. Some surveys and some papers addressing issues such as convergence and acceleration. RWN

Numerical Analysis, T(18: 1, 2), S, P*. Spectral Approximation of Linear Operators. Francoise Chatelin. Comp. Sci. & Appl. Math. Academic Pr, 1983, xix + 458 pp, $69.50. [ISBN: 0-12-170620-6] A thorough and up-to-date exposition of the application of functional analysis applied to eigenvalue problems for linear operators. Emphasizes convergence of operators, spectral approximation of non-self-adjoint generators, use of perturbation theory, and computable error bounds. Includes examples, applications, reviews of eigenvalue problems, and functional analysis, exercises and solutions and an extensive bibliography. RWN

Numerical Analysis, P. Approximate Solution of Plastic Flow Theory Problems. Vadim G. Korneev, Ulrich Langer. Teubner-Texte zur Math., B. 69. BG Teubner, 1984, 252 pp, 26M (P). Basis in elasticity and the flow theory in plasticity. Existence, uniqueness and regularity results. Incremental loading and finite element methods and their convergence. RWN

Numerical Analysis, P. Lecture Notes in Mathematics-1105: Rational Approximation and Interpolation. Ed: P.R. Graves-Morris, E.B. Saff, R.S. Varga. Springer-Verlag, 1984, xii + 528 pp, $25.50 (P). [ISBN: 0-387-13899-4] 43 refereed papers from a joint U.K.-U.S. conference. Includes surveys, theory, numerical methods, and applications. RWN

Numerical Analysis, T(15-16: 1), S, L. An Introduction to the Numerical Solution of Differential Equations. Douglas Quinney. EEE Res. Stud., V. 3. Wiley, 1985, xi + 283 pp, $34.95. [ISBN: 0-471-90849-5] A wonderful text for a first course in the numerical solution of differential equations. Covers recurrence relations, ordinary differential equations, and hyperbolic, parabolic and elliptical partial differential equations. SM

Numerical Analysis, P. Fourier Analysis of Numerical Approximations of Hyperbolic Equations. Robert Vichnevetsky, John B. Bowles. Stud. in Appl. Math. SIAM, 1982, xi + 140 pp, $21.50. [ISBN: 0-89871-181-9] On the use of Fourier analysis to study the behavior of numerical solutions to hyperbolic equations. Includes wave theory, group velocity analysis and time-Fourier analysis. RWN

Numerical Analysis, T(15-18: 1, 2), P, L. Computational Methods for Integral Equations. L.M. Delves, J.L. Mohamed. Cambridge U Pr, 1985, xii + 376 pp, $69.50. [ISBN: 0-521-26629-7] Primarily a reference text for numerical methods applied to one-dimensional integral equations. Concentrates on quadrature and expansion methods for Fredholm equations of the second kind, but includes other items also. Deals with numerical performance of algorithms as a major concern. BC

Numerical Analysis, T(16-17: 1). Numerical Solution of Partial Differential Equations: Finite Difference Methods, Third Edition. G.D. Smith. Appl. Math. & Comp. Sci. Ser. Clarendon Pr, 1985, xiii + 337 pp, $19.95. [ISBN: 0-19-859641-3] Covers the standard topics in finite difference methods for parabolic, hyperbolic, and elliptic equations. This edition features an extended treatment of stability theory and an improved presentation of iterative methods. (First Edition, TR, December 1976; Second Edition, TR, August-September 1979.) AO

Functional Analysis, P. Selfadjoint Operators in Spaces of Functions of Infinitely Many Variables. Yu. M. Berezanskiĭ. Transl. of Math. Mono., V. 63. AMS, 1986, xv + 383 pp, $109. [ISBN: 0-8218-4515-2] The author constructs spaces of functions of infinitely many variables as infinite tensor products of spaces of single variable functions and the projective limits of such spaces. He then studies questions in the spectral theory of selfadjoint and normal operators acting on spaces of these functions. These topics have applications in quantum field theory and the study of random processes in probability. AM

Functional Analysis, S(18), P. Treatise on the Shift Operator: Spectral Function Theory. N.K. Nikol'skiĭ. Transl: Jaak Peetre. Grund. der math. Wissenschaften, B. 273. Springer-Verlag, 1986, xi + 491 pp, $64.50. [ISBN: 0-387-15021-8] An introduction to non-classical spectral theory focused on the corona theorem of Carleson and connections between function theory and spectral theory. Revised and enlarged version of original Russian edition, including additional appendix on singular numbers of Hankel operators. Large bibliography. BH

Functional Analysis, T(17-18: 2), S, P. Lineare Funktionalanalysis: Eine anwendungsorientierte Einführung. Hans Wilhelm Alt. Springer-Verlag, 1985, ix + 292 pp, DM 34 (P). [ISBN: 0-387-15280-6] A terse, modern introductory text. Exercises, most of them with solutions. JD-B

Analysis, S(18), P. Bernstein Polynomials, Second Edition. G.G. Lorentz. Chelsea, 1986, x + 134 pp, $13.95. [ISBN: 0-8284-0323-6] Second Edition of 1953 text with an appendix listing important

papers in the field since 1953. First three (of four) chapters "can be considered an introduction to a theory of singular integrals by means of Bernstein polynomials." Topics include moment problems, Hausdorff methods of summation, rearrangements of functions, and analytic functions. BH

Analysis, T(14-15: 1, 2), L. Analysis: An Introduction to Proof. Steven R. Lay. Prentice-Hall, 1986, x + 285 pp, $31.95. [ISBN: 0-13-032996-7] Covers the basic topics of advanced calculus, paying careful attention to techniques of proof. (Two chapters, Logic and Proof, and Sets and Functions, comprise nearly a third of the text.) Lots of exercises, many hints provided. BC

Analysis, P. Lecture Notes in Mathematics-1085: Asymptotics of Analytic Difference Equations. Geertrui K. Immink. Springer-Verlag, 1984, v + 134 pp, $7.50 (P). [ISBN: 0-387-13867-6] Asymptotic solutions for equations with asymptotically represented operators. Linear and nonlinear difference equations. Existence theory for right inverses of difference operators. Employs Gevrey classes. Applies to block triangularzation problems. RWN

Analysis, T, S(17-18), P, L. Algebraic Theory of Measure and Integration, Second English Edition. C. Caratheodory. Chelsea, 1986, 378 pp, $19.95. A systematic and unified exposition of a very general theory of measure, based on the analogue, on Boolean rings, of ordinary point functions. A reprint of the 1963 English edition. LCL

Algebraic Geometry, P. Cubic Forms: Algebra, Geometry, Arithmetic, Second Edition. Yu. I. Manin. Transl: M. Hazewinkel. Math. Lib., V. 4. Elsevier Science, 1986, x + 326 pp, $64.75. [ISBN: 0-444-87823-8] Can the chord-and-tangent method for cubic (i.e., elliptic) curves be mimicked for cubic surfaces? Much algebraic geometry is needed just to make sense of the qeustion. Original 1974 text, with new appendix to report on recent progress. (First Edition, TR, February 1975.) BC

Differential Geometry, P. Complex Differential Geometry and Nonlinear Differential Equations. Ed: Yum-Tong Siu. Contemp. Math., V. 49. AMS, 1984, xiv + 184 pp, $21 (P). [ISBN: 0-8218-5049-0] Thirteen papers from a conference at Bowdoin College, Maine, in 1984. Primarily for specialists. BC

Differential Geometry, S(17-18), P. Geometric Theory of Foliations. César Camacho, Alcides Lins Neto. Transl: Sue E. Goodman. Birkhauser Boston, 1985, 205 pp, $42. [ISBN: 0-8176-3139-9] A foliation is a decomposition of a manifold into a union of connected disjoint submanifolds called leaves, which pile up locally like pages of a book. This monograph presents an introduction to foliations and discusses the connection between the theory of foliations and differential topology; also contains applications in dynamical systems. AM

Differential Geometry, P. Lecture Notes in Physics-244: Selected Topics in Gauge Theories. Walter Dittrich, Martin Reuter. Springer-Verlag, 1986, 315 pp, $18.60 (P). [ISBN: 0-387-16064-7] A collection of lectures and seminar talks on recent developments in field theory. Not intended as an introduction, but aimed at the reader with a basic knowledge of quantum field theory. AM

Differential Geometry, P*. Spinors and Space-Time, Volume 2: Spinor and Twistor Methods in Space-Time Geometry. Roger Penrose, Wolfgang Rindler. Cambridge U Pr, 1986, ix + 501 pp, $89.50. [ISBN: 0-521-25267-9] The twistor program establishes a relationship between Minkowski space time and complex three-dimensional projective space. Points in projective space correspond to null rays in Minkowski space. Under this transformation, the equations of mathematical physics may be studied in terms of the complex structure of projective space. The book introduces the theory of twistors and studies applications of twistors and 2-spinors to the study of space time. AM

Differential Geometry, P. Manifolds of Nonpositive Curvature. Werner Ballmann, Mikhael Gromov, Viktor Schroeder. Progress in Math., V. 61. Birkhauser Boston, 1985, iv + 263 pp, $37. [ISBN: 0-8176-3181-X] Let V and V* be compact locally symetric spaces of non-positive curvature. The Mostow Rigidity Theorem shows that, given certain hypotheses, if the fundamental groups of V and V* are isomorphic, then V and V* are isomorphic. This book presents a generalization of this theorem as well as the necessary background material in the area of non-positively curved manifolds. AM

Geometry, T(16-17: 1, 2), P, L. Geometric Modeling. Michael E. Mortenson. Wiley, 1985, xvi + 763 pp, $36.95. [ISBN: 0-471-88279-8] Covers much of the mathematics underlying modern computer-aided design and computer-aided manufacturing systems. Organized into three main sections: parametric geometry, solid modeling, and applications. AO

Geometry, S(15-17), P, L*. The Mathematical Description of Shape and Form. É.A. Lord, C.B. Wilson. Ser. in Math. & Its Applic. Halsted Pr, 1984, 260 pp. [ISBN: 0-470-20043-X] A unique survey of shapes and patterns backed up by mathematical representations from linear algebra and vector calculus: topological and analytic mappings, singularities, symmetry, tessellations, minimal surfaces, catastrophe surfaces, space-filling curves, approximations, projections. The authors, formerly a mathematician and a physicist, are now professors of architectural science. LAS

Geometry, T(16: 1), S, P, L. Projective Geometry and Its Applications to Computer Graphics. Michael A. Penna, Richard R. Patterson. Prentice-Hall, 1986, xi + 403 pp, $39.95. [ISBN: 0-13-730649-0] Includes an introduction to perspective and projective geometry, analytic projective geometry and applications of this material to problems in computer graphics (both two and three-dimensional and vision systems). Prerequisites: first course in computer programming, analytic geometry and matrix theory. JNC

Algebraic Topology, P. Lecture Notes in Mathematics-1176: H_∞ Ring Spectra and their Applications. R.R. Bruner, et al. Springer-Verlag, 1986, vii + 388 pp, $28.80 (P). [ISBN: 0-387-16434-0] Spectra with enriched multiplicative structure abound in homotopy and homology theory. Previous ad hoc applications are re-derived here in the larger context of extended powers of spectra. The unpleasant details are deferred to a second (forthcoming) volume. BC

Topology, P. Continuous Lattices and Their Applications. Ed: Rudolf-E. Hoffmann, Karl H. Hofmann. Lect. Notes in Pure & Appl. Math., V. 101. Dekker, 1985, x + 369 pp, $69.75 (P). [ISBN: 0-8247-7331-4] The proceedings of a conference on categorical and topological aspects of continuous lattices held at the University of Breman, West Germany, July 2-3, 1982. Note price! LCL

Optimization, P. Lecture Notes in Economics and Mathematical Systems-255: Nondifferentiable Optimization: Motivations and Applications. Ed: V.F. Demyanov, D. Pallaschke. Springer-Verlag, 1985, vi + 349 pp, $25.80 (P). [ISBN: 0-387-15979-7] Proceedings of the International Institute for Applied System Analysis Workshop on Nondifferential Optimization, Sopron, Hungary, September 17-22, 1984. Presents recent research on nonsmooth analysis, multicriteria optimization, algorithms and applications. SM

Optimization, P. Topics in Relaxation and Ellipsoidal Methods. M. Akgül. Research Notes in Math., No. 97. Pitman, 1984, 322 pp, $24.95 (P). [ISBN: 0-273-08634-0] A study of the complexity of a number of recent algorithms for the exact and for the approximate solution to linear and to convex programming problems. RWN

Probability, P. Limit Theorems for Sums of Random Variables. Ed: A.A. Borovkov. Adv. in Prob. Theory. Optimization Software, 1985, xii + 301 pp, $64. [ISBN: 0-911575-17-0] A collection of nine papers by Soviet probabilists. Most cover convergence rates or asymptotic behavior for specific stochastic estimates. TAV

Probability, P. Lecture Notes in Mathematics-1158: Stochastic Processes--Mathematics and Physics. Ed: S. Albeverio, Ph. Blanchard, L. Streit. Springer-Verlag, 1986, vi + 257 pp, $21.30 (P). [ISBN: 0-387-15998-3] Proceedings of the first in a series of symposia organized by the Bielefeld-Bochum Research Center for Stochastic Study (Germany). The seventeen papers vary from purely theoretical mathematics to applied physics. TAV

Probability, P. Contiguity and the Statistical Invariance Principle. P.E. Greenwood, A.N. Shiryayev. Stoch. Mono., V. 1. Gordon & Breach, 1985, viii + 236 pp, $39. [ISBN: 2-88124-013-5] First volume of a new series, published in association with the journal Stochastics, dealing with stochastic processes and their applications. Series will cover material in a more extended and expository form than is possible in the journal. First part investigates the notion of contiguity (an asymptotic version of the idea of absolute continuity) and the related notion of asymptotic entire separation. Second part concerns the statistical invariance principle (functional central limit theorem) and recent extensions. RSK

Statistics, T*(16: 1), L. Time Series Analysis. Jonathan D. Cryer. Duxbury Pr, 1986, xi + 286 pp. [ISBN: 0-87150-963-6] Presents theory and applications of time domain ARIMA (Box-Jenkins) models at a level accessible to a wide variety of students. Extensively illustrated with real data analyzed using the Minitab statistical system. Theory presumes a calculus-based introduction to statistics. RSK

Statistics, T(15-16: 1), P. Introduction to Statistical Quality Control. Douglas C. Montgomery. Wiley, 1985, xviii + 520 pp, $36.95. [ISBN: 0-471-80870-9] Engineering-oriented text, divided into three main parts. First part briefly presents statistical methods from the quality-assurance point of view. Second and main part (over half the text) covers statistical process control, primarily using control charts. Final part deals with acceptance sampling procedures. RSK

Statistics, T(13-14: 1, 2). Statistics: Concepts and Applications. David R. Anderson, Dennis J. Sweeney, Thomas A. Williams. West, 1986, xxvi + 687 pp, $33.95. [ISBN: 0-314-93146-5] Presupposes a course in college algebra. Many examples of real applications. The usual topics plus some exploratory data analysis, optional discussion of use of MINITAB, contingency tables, and multiple regression. FLW

Statistics, T(16-17: 1), S*, P*, L*. Plots, Transformations, and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis. A.C. Atkinson. Oxford Stat. Sci. Ser. Clarendon Pr, 1985, xiii + 282 pp, $45. [ISBN: 0-19-853359-4] A valuable book to use while carrying out regression analysis with a computer package. Considers checking model assumptions, outliers, transformation of data, generalized linear models, and robust estimation. FLW

Statistics, T(13-14: 1). General Statistics. Warren Chase, Fred Bown. Wiley, 1986, xiv + 635 pp, $27.95. [ISBN: 0-471-86862-0] Presupposes only high school algebra. The usual topics plus optional sections using MINITAB output. There is a supplement available that treats both MINITAB and SPSS. Includes contingency tables and some nonparametric tests. FLW

Statistics, S(15-17), L. Counterexamples in Probability and Statistics. Joseph P. Romano, Andrew F. Siegel. Stat./Prob. Ser. Wadsworth, 1986, xxiv + 303 pp, $34.95. [ISBN: 0-534-05568-0] Collection of counterexamples grouped by chapters in ten general areas, from probability spaces to hypothesis testing. Each chapter begins with an introduction which includes pertinent definitions and results. An appendix contains references to some further examples. RSK

Statistics, P*. Statistical Inference in Linear Models: Statistical Methods of Model Building, Volume I. Ed: Helga and Olaf Bunke. Transl: John Bibby, Michal Basch. Wiley, 1986, 614 pp, $72.95. [ISBN: 0-471-10334-9] In the Wiley Series in Probability and Mathematical Statistics, Applied Section. Translation with slight modifications of the first volume of a German trilogy by K.M.S. Humak (an acronym for a group of German statisticians headed by Helga and Olaf Bunke), published in 1977. Presents a comprehensive, unified and rigorous development of the theory of the linear model in the areas of estimation, testing, confidence regions, Bayesian methods, and optimal design. RSK

Statistics, T*(13: 1, 2). Statistics: The Exploration and Analysis of Data. Jay Devore, Roxy Peck. West, 1986, xvii + 699 pp, $33.95. [ISBN: 0-314-93172-4] Modern treatment of elementary statistics. Emphasizes concepts and an intuitive understanding of core methodology. Makes excellent use of real data in both examples and problems. RSK

Statistics, T(15-16: 1). Probability and Statistical Inference, Volume 2: Statistical Inference, Second Edition. J.G. Kalbfleisch. Texts in Stat. Springer-Verlag, 1985, xiii + 360 pp, $29.80. [ISBN: 0-387-96183-6] Extensive revision and reorganization (this edition is also hardbound and typeset) of the second volume of the author's 1979 two-volume text (TR, March 1980). Somewhat non-standard treatment, emphasizing probability models and general techniques for deriving estimates and tests from the likelihood function rather than presenting specific tests and procedures. RSK

Statistics, T(17-18: 2), P. Modern Multivariate Statistical Analysis: A Graduate Course and Handbook. Minoru Siotani, Takesi Hayakawa, Yasunori Fujikoshi. Ser. in Math. & Management Sci., V. 9. American Sciences Pr, 1985, xiv + 759 pp, $39.50 (P). [ISBN: 0-935950-06-0] Comprehensive treatment assuming a background of multivariate methods, mathematical statistics and linear algebra. First half deals with multivariate analogs of univariate methods and techniques. Second half treats topics peculiar to multivariate analysis, including discriminant analysis, principal component analysis, canonical correlation analysis, and methods for selection of variables. Appendices give useful results on matrix theory for multivariate analysis and some specific Fortran programs. RSK

Statistics, T(18), P. Random Data: Analysis and Measurement Procedures, Second Edition (Revised and Expanded). Julius S. Bendat, Allan G. Piersol. Wiley, 1986, xvii + 566 pp. [ISBN: 0-471-04000-2] Updated version of the authors' 1971 First Edition (TR, August-September 1974) to "reflect recent changes in model formulations, statistical error evaluations, data collection procedures, and computational algorithms." Designed to be an analytical companion to the authors' 1980 applications-oriented book Engineering Applications of Correlation and Spectral Analysis (TR, March 1981). RSK

Statistics, P. Sequential Tests. Karl-Heinz Eger. Teubner-Texte zur Math., B. 74. BG Teubner, 1985, 172 pp, 17,50 M (P). Theoretical monograph dealing primarily with properties of sequential likelihood ratio tests and a method for the computation of their characteristics. RSK

Statistics, P. Order Dependence. B.F. Schriever. CWI Tract, V. 20. Math Centrum, 1986, iii + 115 pp, Dfl. 17.60 (P). [ISBN: 90-6196-294-3] Theoretical monograph dealing with some aspects of the analysis of ordered contingency tables, mainly bivariate, where the order is usually induced rather than natural. Approach is non-parametric rather than log-linear, and is motivated by the technique of correspondence analysis. RSK

Computer Literacy, L. Computing Information Directory: A Comprehensive Guide to the Computing Liturature. Ed: Darlene Myers Hildebrandt. Pedaro, 1985, v + 557 pp, (P). [ISBN: 0-933133-00-5] A listing of various information sources which is difficult to describe because of the inherent disorganization of its topic. Various chapters provide lists of journals (all kinds), university computer center newsletters, languages, books, technical reports as well as software reviews and hardware sources. Reasonably comprehensive, mildly eclectic (neither C nor Lisp occurs in the short list of computer languages), but probably a helpful place to start looking. This 1985 edition is a successor to Computer Science Resources, 1981. JAS

Computer Programming, T(15-16: 1), S, L. MODULA-2: A Software Development Approach. Gary Ford, Richard Wiener. Wiley, 1985, xvi + 400 pp, $24.45 (P). [ISBN: 0-471-87834-0] An introduction to Modula-2 and modern software development methodology. Presumes prior experience with Pascal. A good textbook for an advanced programming course. AO

Reviewers

DFA: David F. Appleyard, Carleton; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; GF: Giovanna Fjelstad, St. Olaf; SG: Steven Galovich, Carleton; BH: Bruce Hanson, St. Olaf; PH: Paul Humke, St. Olaf; KK: Kenneth Kaminsky, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; AM: Alan Magnuson, St. Olaf; SM: Steve McKelvey, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

CENTER SECTION INDEX

Volume 93, 1986

Primality table for $a = 1$

2^t	p_t	$p_t^2 - 1$
32	193	37248
64	607	368448
128	1747	3052008
256	11261	1 26810120
512	21911	4 80091920
1024	100417	100 83573888

Any factor less than p_t^2 is prime.

References

1. R. P. Brent, An improved Monte Carlo factorization algorithm, BIT, 20(1980) 176–184.
2. John Brillhart, D. H. Lehmer, and J. L. Selfridge, New primality criteria and factorizations of $2^m \pm 1$, Math. Comp., 29(1975) 620–647.
3. J. M. Pollard, A Monte-Carlo method for factorization, BIT, 15(1975) 331–334.
4. Carl Pomerance, J. L. Selfridge, and Samuel S. Wagstaff, Jr., The Pseudoprimes to $25 \cdot 10^9$, Math. Comp., 35(1980) 1003–1026. (The $30 prize on page 1025 has been upped to $120.)
5. J. L. Selfridge and Richard K. Guy, Primality Testing with Application to Small Machines, Proc. WSU Conf. Number Theory, Pullman, WA, 1971, 45–51.

THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

For instructions about submitting material for publication in this department see the inside front cover.

A "GREAT THEOREMS" COURSE IN MATHEMATICS

WILLIAM DUNHAM
Department of Mathematics, Hanover College, Hanover, IN 47243

I recently came across an advertisement for a new textbook on number theory. The ad's main thrust was to acclaim the applications of number theory in the construction of unbreakable codes. In essence, it implied that such a real-world application had at last legitimized number theory as a respectable area of human inquiry.

As I tossed the advertisement aside, I was left feeling decidedly uneasy. Even number theory —long regarded as the purest of the pure—must now be marketed as applied mathematics, somehow vital to the national security, in order to be considered worthwhile. Of course, it was not its strategic utility that led Euclid or Fermat or Euler or Gauss to devote so much of their energy and genius to the "higher arithmetic." These gentlemen did not feel compelled to justify their number theoretic work by its real-world applications, any more than Shakespeare had to apologize for writing love sonnets instead of cookbooks or than Van Gogh had to apologize for painting canvases instead of billboards.

Of course, applications are inseparable, both logically and historically, from mathematics itself, and recent advances in applied number theory are surely interesting. Yet I am afraid that the mathematics professor of today tends to ignore the centuries-old conception of mathematics as an artistic exercise in pure reason that can provide fascinating glimpses into the logical relationships of number and space. Too often, we fall into the easy trap of stressing only the applications of mathematics to commerce or computers. In so doing, I think we ignore an

important mathematical tradition dating back to classical times; and we run the risk of undercutting our liberal arts colleagues in literature or history or music, whose disciplines can not promise their students lucrative job offers after graduation.

Two years ago, with such thoughts in mind, I submitted a grant proposal to the Lilly Endowment, Inc., and was fortunate enough to get support in creating a new course I call "The Great Theorems of Mathematics." My idea was to examine a score or so of the most ingenious theorems I know, stressing the historical epoch in which they appeared and the biographies of their creators, but mainly presenting their PROOFS so that the student could behold the workings of the most fertile minds of our discipline. Clearly, this course would be all but useless in any vocational sense. But so are courses in classical history or modern fiction or any of a number of staples of the liberal arts curriculum.

By focusing the course upon a handful of proofs, I hoped to distinguish it from the typical offering in math history which, in trying to mention everyone from Thales to Bourbaki, has little time to delve into the mathematics itself. This approach of talking *about* great mathematics without doing it strikes me as analogous to studying art history in a dark room—there, in the pitch blackness, the professor might praise the talents of Leonardo da Vinci and describe in words the particularly beautiful portrait of a dark-haired young woman with crossed hands and a mysterious smile on her face, but the poor students, sitting in the dark, would surely be unable to appreciate the genius of the artist without seeing the painting for themselves.

By contrast, I wanted to teach a course comparable to a literature offering on "Landmarks of the Novel" or a music course on "The Great Symphonies," where the novels are actually READ or the music actually HEARD. The only difference was that I would be examining "the Great Theorems," the creative milestones of mathematics.

Obviously, my crucial task in designing this course was to choose the theorems themselves. Here I adopted five criteria of selection:

1. The theorems had to be understandable to a student with only a calculus background. Were I to accept anything less than calculus proficiency, I would be handicapped by a lack of mathematical sophistication on the part of the student; but, were I to demand more extensive pre-requisites, I would condemn myself to teaching only mathematics majors, and few of them at that. So, while de Branges' recent proof of the Bieberbach Conjecture may qualify as a "great theorem," no undergraduate student would have a prayer of understanding it.

2. The chosen theorems had to be important. Not just clever deductions, they should be results with implications that reverberated through the history of mathematics and both answered, and raised, crucial questions. This criterion effectively eliminated puzzles, tricks, and brain-teasers, regardless of their cleverness. To devote time to such trivialities would be much like taking time from a "Great Music" course to study the jingles of television commercials.

3. It was essential that I sample the work of historically important mathematicians. I thus had to include something from Euclid, from Archimedes, from Newton, from Euler. No course examining the milestones of mathematics would dare omit these giants.

4. I wanted theorems that spanned the centuries and theorems that covered many subdisciplines of mathematics. I thereby hoped to impart an appreciation for the sweep of mathematics as well as the sweep of history.

5. Most importantly, I insisted upon choosing theorems whose proofs were ingenious. The arguments I examined should exhibit an intellectual sparkle. The proofs might take odd twists or unexpected turns yet would achieve their object with brilliance and verve. In short, I was looking for masterpieces—the "Mona Lisas" or "Hamlets" of mathematics.

I thus generated my list of great theorems, around whose proofs I wove the historical and biographical components of the course. Obviously, it was imperative to retain as much of the

original logical strategy of the proofs as possible, although occasionally an archaic notation or approach had to be replaced by its modern counterpart. Just as obviously, the matter of personal, I would even say aesthetic, taste entered into the final selection of topics.

Briefly, let me mention the theorems I selected:

I begin with Hippocrates' Quadrature of the Lune from 440 B.C. ([**5**], p. 185), a short but interesting geometric argument, not to mention the oldest surviving mathematical proof, and one that raised questions about the quadrature of the circle that would echo through mathematics for centuries.

Euclid is represented in my course by three theorems, all taken verbatim from the *Elements*. Euclid's proof of the Pythagorean Theorem ([**6**], Proposition I.47), a genuine classic for two millenia, requires a few days of preliminary work on the early structure of the *Elements* and generates a host of assignments exploring alternative proofs of this famous result. Our other two Euclidean propositions are the converse of the Pythagorean Theorem ([**6**], Proposition I.48) and the demonstration that no finite collection of positive integers can exhaust all the primes ([**6**], Proposition IX.20); of this latter result, the British mathematician G. H. Hardy wrote "...(it is) as fresh and significant as when it was discovered—two thousand years have not written a wrinkle on (it)." ([**4**], p. 92).

We look at two theorems from Archimedes—his "double reductio ad absurdum" proof from *Measurement of a Circle* ([**7**], Proposition I), relating a circle's area and circumference, which in modern terms becomes the familiar formula $A = \pi r^2$; and his remarkable theorem on the surface area of a sphere from *On the Sphere and the Cylinder* ([**7**], Proposition 33) which gives us $S = 4\pi r^2$.

A final result from classical times is Ptolemy's construction of a table of chords from the *Almagest* ([**8**], Book I, Chapter 10). This not only features a wonderful logical development but also suggests an assignment in which students create their own tables by the post-classical technique of Taylor Series expansions.

I next look at Cardano's solution of the cubic equation from the 16th Century ([**9**], p. 203–206). This clever argument, coupled with Cardano's rise from humble origins to international notoriety, makes for a "Horatio Algebra" story if ever there was one. We then venture into 17th Century England to examine Newton in action, applying his binomial theorem in an ingenious estimate of the value of π ([**11**], p. 223–227) and in the original description of what we now call "Newton's Method" ([**12**], p. 219–223).

Returning to the Continent, we take up Johann Bernoulli's interesting but little known proof of the divergence of the harmonic series ([**10**], p. 321–322), then move on to three theorems of the incomparable Leonhard Euler—his summation of the infinite series of the reciprocals of the squares of the integers ([**2**], p. 83–85), his inductive proof of the Little Fermat Theorem ([**3**], p. 65–67), and his marvelous deduction by which he discovered that, contrary to previous belief, $2^{2^n} + 1$ need not always be prime ([**3**], p. 68–74). (In this last argument, by the way, Euler *used* the Little Fermat Theorem, thereby illustrating the old saying that a little Fermat goes a long way.)

The course concludes in the 19th Century with an examination of Cantor's second proof, via the diagonal process, of the non-denumerability of the continuum ([**13**], p. 278–281) as well as the result now known as "Cantor's Theorem" ([**1**], p. 165–167) and its mind-boggling implications.

I think a course like this is entirely appropriate in today's mathematics curriculum. Admittedly, students do not acquire specific mathematical skills here—they do not learn how to invert matrices or solve differential equations. By the same token, students in an art history course are not trained to paint landscapes or sketch portraits. Like the art historian, my objective is not the acquisition of skills but the appreciation of greatness.

I am pleased to report that "The Great Theorems of Mathematics" has drawn students not only from among math majors and minors but also from those in pre-Law, in the sciences, in philosophy, and especially from among those liberal artists who see their education not as a ticket

to a high-paying career in industry but as the opportunity to match wits with the likes of Archimedes or Isaac Newton.

Finally, I am equally pleased to report that the students have enjoyed the course. And, for those who succeed in understanding these great theorems, I believe there emerges not only a sense of awe in recognizing the genius of others but also a degree of personal satisfaction that one can, indeed, comprehend the works of a master.

References

1. J. Dauben, Georg Cantor: His Mathematics and Philosophy of the Infinite, Harvard Univ. Press, 1979.
2. C. Boehm and G. Faber (eds.), Euler's Opera Omnia (1), Vol. 14, Leipzig, 1925.
3. F. Rudio (ed.), Euler's Opera Omnia (1), Vol. 2, Leipzig, 1915.
4. G. H. Hardy, A Mathematician's Apology, Cambridge Univ. Press, 1969.
5. T. L. Heath, A History of Greek Mathematics, Vol. 1, Dover (reprint), New York, 1981.
6. T. L. Heath (ed.), The Thirteen Books of Euclid's Elements (in 3 Volumes), Dover (reprint), New York, 1956.
7. ______, The Works of Archimedes, Dover (reprint), New York, 1953.
8. R. M. Hutchins (ed. in chief), Great Books of the Western World, Vol. 16, Encyclopedia Britannica, 1952.
9. D. E. Smith (ed.), A Source Book in Mathematics, McGraw-Hill, New York, 1929.
10. D. Struik (ed.), A Source Book in Mathematics: 1200–1800, Harvard Univ. Press, 1969.
11. D. Whiteside (ed.), The Mathematical Papers of Isaac Newton, Vol. III (1670–1673), Cambridge Univ. Press, 1970.
12. ______, The Mathematical Papers of Isaac Newton, Vol. II (1667–1670), Cambridge Univ. Press, 1969.
13. E. Zermelo (ed.), Cantor's Gesammelte Abhandlungen, Berlin, 1962.

PROBLEMS AND SOLUTIONS

*For instructions about submitting **proposed** problems for publication in this department see the inside front cover. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the address given on the inside front cover.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

For instructions about submitting solutions of these Elementary Problems, which should be mailed by April 30, 1987, see the inside front cover. Please place the solver's name and mailing address on each (doubled-spaced) sheet. Include a self-addressed card or label (for acknowledgement).

E 3177. *Proposed by Jordi Dou, Barcelona, Spain.*

Let A, B, C be three points on a circle. Let A_1 (respectively, B_1, C_1) be the intersection of the

tangent line at A (respectively, B, C) with the line through BC (respectively, CA, AB). Prove that the circles ABB_1, BCC_1, CAA_1 and the line $A_1B_1C_1$ have a common point.

E 3178. *Proposed by G. A. Hively, Lockheed Palo Alto Research Laboratory, CA.*

Let x_i and y_j be arbitrary real numbers. Show that the $n \times n$ matrix

$$S = \left(\sin(x_i + y_j)\right)$$

is singular if $n \geqslant 3$.

E 3179. *Proposed by N. J. Lord, Tonbridge School, England.*

Given a real number x, let $f_n(x)$ denote the distance from x to the nearest rational number with denominator n (not necessarily in its lowest form). For which values of x does the series $\sum_{n=1}^{\infty} f_n(x)$ converge?

E 3180. *Proposed by Murray S. Klamkin, University of Alberta, Canada.*

If A, B, C are angles of a triangle, prove that

$$\cos\frac{A}{2} + \cos\frac{B}{2} + \cos\frac{C}{2} \geqslant 1 + \sin\frac{A}{2} + \sin\frac{B}{2} + \sin\frac{C}{2}.$$

E 3181. *Proposed by Zalman Rubinstein, University of Colorado.*

Let $f(z) = z^2 + z$. Construct a cyclic sequence for $f(z)$, that is, a sequence $\{z_n\}_{-\infty}^{\infty}$ of non-zero complex numbers with $z_{n+1} = f(z_n)$ for all integers n and such that

$$\lim_{n \to \infty} z_n = \lim_{n \to -\infty} z_n = 0.$$

E 3182*. *Proposed by Louis Funar, University of Craiova, Romania.*

Let $\mathscr{S}$ be a family of parallel rectangles (i.e., having parallel sides) whose union covers a parallel rectangle of area 4. Prove or disprove that there exists a subset of $\mathscr{S}$ with disjoint interiors whose union has area of at least 1. Extend to $\mathbb{R}^n$.

SOLUTIONS OF ELEMENTARY PROBLEMS

$f_{k+1}(n) = \sum_{d|n} f_k(d)$ and Identically Zero Arithmetic Functions

E 2957 [1982, 498]. *Proposed by Stephen McAdam, The University of Texas at Austin.*

Let f be a function defined on the positive integers. Let $f_1 = f$ and inductively define $f_{k+1}(n) = \sum_{d|n} f_k(d)$ over positive divisors d of n. This operation is well known in elementary number theory. Show that if $k > 1$ and if $f_k = f_1$, then f is identically zero.

Solution I *by Nathaniel Grossman, University of California, Los Angeles.* Suppose

$$F_k(s) = \sum_{n=1}^{+\infty} f_k(n) n^{-s}$$

is the formal Dirichlet series associated with f_k, and set

$$\zeta(s) = \sum_{n=1}^{+\infty} n^{-s}.$$

Then $F_{r+1}(s) = [\zeta(s)]^r F_1(s)$.

Let $f_k = f_1$ with $k > 1$. If f_1 is not identically zero, then F_1 is not the zero element of the ring of formal Dirichlet series. That ring has no divisors of zero, so that $[\zeta(s)]^{k-1} = 1$, the unit of the ring. Both sides are analytic functions of s in the half-plane $\mathrm{Re}(s) > 1$, so the consequence is that the Riemann Zeta Function is constant. This is well known to be false, so the only possibility is that F_1 is the zero element, whence f_1 is identically zero.

Solution II *by Vania D. Mascioni* (*student*), *Swiss Federal Institute of Technology*, *Zurich*. We prove the following generalization:

Define the f_i as in the problem statement and let f_1 be not identically zero. Then the infinite set $\{f_1, f_2, \dots, f_s, \dots\}$ is linearly independent over $\mathbb{C}$.

Proof. Let $s \geq 1$ and α_i, $1 \leq i \leq s$, be complex numbers not all zero, such that $h := \alpha_1 f_1 + \cdots + \alpha_s f_s = 0$, with f_1 not identically zero. If "$*$" denotes the Dirichlet convolution of arithmetical functions and u is defined by $u(n) = 1$, for all n, and I by $I(n) = \delta_{1n}$, we can write $f_{k+1} = u * f_k$. Hence

$$h = \left(\alpha_1 I + \alpha_2 u + \cdots + \alpha_s u^{s-1}\right) * f_1 \equiv 0$$

(where $u^k := u * u^{k-1}$). Since the ring $(\mathbb{C}^{\mathbb{N}}, +, *)$ of the arithmetical functions has no zero divisors, it must be

$$\bar{h} := \alpha_1 I + \cdots + \alpha_s u^{s-1} \equiv 0.$$

Consider now $\bar{h}(1), \bar{h}(p_1), \bar{h}(p_1 p_2), \dots, \bar{h}(p_1 \cdots p_{s-1})$, where $p_1, \dots, p_{s-1}$ are different primes. Since u is multiplicative, we have

$$u^k(p_1 \cdots p_{s-1}) = u^k(p_1) \cdots u^k(p_{s-1}) = k^{s-1}.$$

We obtain also for $\alpha_1, \dots, \alpha_s$ the homogeneous linear system

$$\begin{aligned}
\alpha_1 + \alpha_2 + \alpha_3 + \cdots + \alpha_s &= 0,\\
\alpha_2 + 2\alpha_3 + \cdots + (s-1)\alpha_s &= 0,\\
&\vdots\\
\alpha_2 + 2^{s-1}\alpha_3 + \cdots + (s-1)^{s-1}\alpha_s &= 0,
\end{aligned}$$

whose matrix is the Vandermonde matrix of order $s - 1$ and hence its determinant is not zero. It follows that $\alpha_1 = \cdots = \alpha_s = 0$, and $\{f_1, \dots, f_s\}$ is linearly independent. Since this is true for all $s \in \mathbb{N}$, the infinite set $\{f_1, \dots, f_s, \dots\}$ is linearly independent over $\mathbb{C}$, as we had to prove.

Also solved by 27 other readers and the proposer.

A Unique Type of Partition of the Positive Integers

E 2977 [1982, 756; 1986, 300]. **Correction to the reference cited.**

F. H. Kierstead, Jr., has pointed out that the problem did appear in the *Journal of Recreational Mathematics*, but not at the location cited. The problem was published in volume 9, number 4, page 298, and the solution in volume 10, number 4, pages 314–315.

An Abundance of Deficient Numbers

E 3002 [1983, 400]. *Proposed by Charles R. Wall, Trident Technical College.*

A number n is called deficient, perfect or abundant if the sum of its divisors is $<$, $=$, or $>$ $2n$. Prove that if the g.c.d. of a and b is deficient, then there exist
(1) infinitely many deficient integers $n \equiv a \pmod{b}$;
(2) infinitely many abundant integers $n \equiv a \pmod{b}$.

Solution by N. J. Fine, Deerfield Beach, Florida. (1) Let $d = (a, b)$, so $\sigma(d) \le 2d - 1$. Let $a = d\alpha$, $b = d\beta$, so $(\alpha, \beta) = 1$. There are infinitely many primes of the form $\alpha + \beta x$. For every such prime $p > 2d$, let $n = dp = a + bx \equiv a \pmod{b}$. Then

$$\frac{\sigma(n)}{2n} = \frac{\sigma(dp)}{2dp} = \frac{\sigma(d)}{2d} \cdot \frac{\sigma(p)}{p} \le \frac{2d-1}{2d} \cdot \frac{p+1}{p}$$
$$= \left(1 - \frac{1}{2d}\right)\left(1 + \frac{1}{p}\right) < \left(1 - \frac{1}{2d}\right)\left(1 + \frac{1}{2d}\right)$$
$$= 1 - \frac{1}{4d^2} < 1.$$

Hence n is deficient.

(2) Let $p_1, p_2, \ldots$ be the primes $> b$. Since $\Sigma 1/p = \infty, \exists k$ such that $1/p_1 + 1/p_2 + \cdots + 1/p_k > 1$. For every $\alpha_1, \alpha_2, \ldots \alpha_k > 0$ we can solve

$$n = a + bx \equiv 0 \left(\bmod p_1^{\alpha_1} \cdots p_k^{\alpha_k}\right).$$

For such n we have

$$\frac{\sigma(n)}{n} \ge \left(1 + \frac{1}{p_1}\right)\left(1 + \frac{1}{p_2}\right) \cdots \left(1 + \frac{1}{p_k}\right) \ge 1 + \sum_{j=1}^{k} \frac{1}{p_j} > 2.$$

Thus n is abundant. Note that we did not use the deficiency of (a, b).

Also solved by R. Breusch, W. Castrellon (Colombia), P. J. Costello, L. Cseh and I. Merenyi (Romania), L. Filep (Hungary), L. L. Foster, D. Jeffords, K. Kearnes, L. Kuipers (Switzerland), D. Moews, R. L. Persky, J. Propp, R. E. Rogers, R. Stong, University of South Alabama Problem Group, and the proposer.

Generating Similar Triangles

E 3009 [1983, 482]. *Proposed by Chris Jantzen, University of Chicago.*

Points X, Y, Z are chosen on the sides of a triangle ABC such that

$$\frac{AX}{XB} = \frac{BY}{YC} = \frac{CZ}{ZA} = k$$

and a triangle PQR is formed using CX, AY and BZ as sides. The operation is repeated on the triangle PQR, that is, the points X', Y', Z' are chosen on the sides of PQR such that

$$\frac{PX'}{X'Q} = \frac{QY'}{Y'R} = \frac{RZ'}{Z'P} = k$$

and a triangle LMN is formed using RX', PY' and QZ' as sides. Show that LMN is similar to ABC and find the ratio of similarity.

Editorial note. The statement of the problem should have included the requirement that triangle PQR has the same orientation as triangle ABC. With this requirement, the second line of ratio equalities should be

$$\frac{PX'}{X'Q} = \frac{QY'}{Y'R} = \frac{RZ'}{Z'P} = \frac{1}{k}.$$

There were two different interpretations of the phrase "... triangle PQR is formed using CX, AY, and BZ as sides", which led to solutions to two different problems. We give a sample of each below.

Solution I *by Howard Eves, University of Maine.* Set $t = k/(1 + k)$. Then

$$AX = tAB, \quad BY = tBC, \quad CZ = tCA.$$

Denoting the sides of triangle ABC respectively opposite A, B, C by a, b, c, and denoting

AY, BZ, CX by p, q, r, we have, by Stewart's Theorem (see, e.g., Howard Eves, *A Survey of Geometry*, *Revised Edition*, p. 58, prob. 13),

$$(AY)^2 BC = (AC)^2 BY + (AB)^2 YC - (BY)(YC)(BC),$$

or

$$p^2 = tb^2 + (1-t)c^2 - t(1-t)a^2.$$

Similarly,

$$q^2 = tc^2 + (1-t)a^2 - t(1-t)b^2,$$
$$r^2 = ta^2 + (1-t)b^2 - t(1-t)c^2.$$

Denote PY', QZ', RX' by l, m, n. Then, again by Stewart's Theorem,

$$\begin{aligned} l^2 &= tr^2 + (1-t)q^2 - t(1-t)p^2 \\ &= t[ta^2 + (1-t)b^2 - t(1-t)c^2] + (1-t)[tc^2 + (1-t)a^2 - t(1-t)b^2] \\ &\quad -t(1-t)[tb^2 + (1-t)c^2 - t(1-t)a^2] \\ &= a^2[t^2 + (1-t)^2 + t^2(1-t)^2] + b^2[t(1-t) - t(1-t)^2 - t^2(1-t)] \\ &\quad +c^2[-t^2(1-t) + t(1-t) - t(1-t)^2] \\ &= (t^4 - 2t^3 + 3t^2 - 2t + 1)a^2 \\ &= (t^2 - t + 1)^2 a^2. \end{aligned}$$

Similarly, $m^2 = (t^2 - t + 1)^2 b^2$, $n^2 = (t^2 - t + 1)^2 c^2$. It follows that triangle LMN is similar to triangle ABC with $s = t^2 - t + 1$ as the ratio of similarity. In terms of k one easily finds

$$s = (k^2 + k + 1)/(1 + k)^2.$$

Solution II *by Jan van de Craats, The Netherlands.* The statement of the problem is ambiguous, since it is not clear whether triangle PQR is, or is not orientated the same as triangle ABC. If the orientations are equal, then the proposition is false: take, for instance, $A = 0, B = 1, C = i$ in the complex plane, and $k = 1/2$. Then, by easy computations,

$$\begin{aligned} &X = 1/3, \quad Y = (2+i)/3, \quad Z = 2i/3, \\ &P = (1/7)A + (6/7)Y = (4+2i)/7, \\ &Q = (1/7)B + (6/7)Z = (1+4i)/7, \\ &R = (1/7)C + (6/7)X = (2+i)/7, \\ &X' = (1/3)Q + (2/3)P = (9+8i)/21, \quad Y' = (4+9i)/21, \quad Z' = (8+4i)/21, \end{aligned}$$

and

$$L = (12 + 20i)/49, \quad M = (17 + 12i)/49, \quad N = (20 + 17i)/49.$$

Clearly, ABC and LMN are not similar. However, if ABC and PQR have opposite orientations, then ABC and LMN are even *homothetic* triangles. Therefore, we restate the problem, taking the opportunity to choose a more symmetric notation.

Restatement of the problem. Let a triangle $A_1A_2A_3$ and a real number $k \neq 0, \pm 1$ be given. For $i = 1, 2, 3$, define X_i, B_i, Y_i, and C_i in the following way: X_i is the point on $A_{i+1}A_{i+2}$ such that $\overline{A_{i+1}X_i} : \overline{X_iA_{i+2}} = k : 1$ (signed distances; indices are taken modulo 3), $B_i = A_{i+1}X_{i+1} \cap A_{i+2}X_{i+2}$, Y_i is the point on $B_{i+1}B_{i+2}$ such that $\overline{B_{i+1}Y_i} : \overline{Y_iB_{i+2}} = 1 : k$, and $C_i = B_{i+1}Y_{i+1} \cap B_{i+2}Y_{i+2}$. Prove that the triangles $A_1A_2A_3$ and $C_1C_2C_3$ are homothetic and find their ratio of similarity.

Solution. We embed the triangle in the complex plane, and define $a_i = A_{i+1} - A_i$. Note that

$$a_1 + a_2 + a_3 = 0. \tag{1}$$

We have $X_i = (A_{i+1} + kA_{i+2})/(k+1)$ and an easy computation yields

$$B_i = \left(A_{i+2} + k^2A_{i+1} + kA_i\right)/\left(k^2+k+1\right).$$

Define $b_i = B_{i+1} - B_i$. Then

$$b_i = \left(a_{i+2} + k^2a_{i+1} + ka_i\right)/\left(k^2+k+1\right). \tag{2}$$

Similarly, if we define $c_i = C_{i+1} - C_i$, then we have

$$\begin{aligned} c_i &= \left(b_{i+2} + k^{-2}b_{i+1} + k^{-1}b_i\right)/\left(k^{-2}+k^{-1}+1\right) \\ &= \left(k^2b_{i+2} + b_{i+1} + kb_i\right)/\left(k^2+k+1\right). \end{aligned}$$

Upon substituting (2) and using (1), we get

$$\begin{aligned} c_i &= \left(\left(k^4+k^2+1\right)a_i + \left(k^3+k^2+k\right)\left(a_{i+1}+a_{i+2}\right)\right)/\left(k^2+k+1\right)^2 \\ &= \left(\left(k^4-k^3-k+1\right)a_i\right)/\left(k^2+k+1\right)^2 \\ &= (k-1)^2\left(k^2+k+1\right)^{-1}a_i. \end{aligned}$$

Thus triangles $C_1C_2C_3$ and $A_1A_2A_3$ are homothetic with ratio of similarity equal to

$$(k-1)^2\left(k^2+k+1\right)^{-1}.$$

Also solved by J. Dou (Spain), R. H. Eddy (Canada), F. Gerrish (England), L. Kuipers (Switzerland), D. Lindsay, O. P. Lossers (The Netherlands), H. J. Ludwig, D. Moews, J. -M. Monier (France), H. S. Morse, I. A. Sakmar (Canada), A. Tissier (France), and the proposer.

A Trivial Group Automorphism

E 3039 [1984, 203]. *Proposed by I. N. Herstein, University of Chicago.*

Suppose that G is a simple nonabelian group. Prove that if ϕ is an automorphism of G such that $x\phi(x) = \phi(x)x$ for all $x \in G$, then $\phi = 1$.

Solution by Thomas J. Laffey, University College Dublin, Ireland. We prove more generally: Let G be a group with no non-trivial *abelian* normal subgroup and suppose ϕ is an automorphism of G such that $x\phi(x) = \phi(x)x$ for all $x \in G$. Then $\phi = 1$.

Fix x in G. The equations

$$x\phi(x) = \phi(x)x, \quad y\phi(y) = \phi(y)y, \quad xy\phi(x)\phi(y) = \phi(x)\phi(y)xy$$

imply

$$y\phi(x)y^{-1} = x^{-1}\phi(x)\cdot\phi(y)x\phi(y)^{-1} \tag{1}$$

for all $y \in G$. Replace x by x^{-1} in (1) and invert to get

$$y\phi(x)y^{-1} = \phi(y)x\phi(y)^{-1}\cdot x^{-1}\phi(x). \tag{2}$$

(1) and (2) imply that $x^{-1}\phi(x)$ commutes with $\phi(y)x\phi(y)^{-1}$ and with $y\phi(x)y^{-1}$ for all $y \in G$. Hence $x^{-1}\phi(x)$ commutes with all elements of N, where N is the *normal* subgroup generated by x, $\phi(x)$ and hence $x^{-1}\phi(x) \in Z(N)$, the centre of N. Since $Z(N)$ is an abelian normal subgroup of G, we have $Z(N) = \{1\}$ and $x = \phi(x)$ as required.

Editorial Note. Several solutions were received containing the assumption that G was finite.

Also solved by A. Bondesen (Denmark), F. Cedó Giné (Spain), R. Guraldo, I. M. Isaacs, T. Jager, O. P. Lossers (The Netherlands), M. L. Newell (Ireland), M. Pettet, A. Shamsuddin (Beirut), D. Spellman, G. L. Wallis, M. B.

Ward, and the proposer. Partially solved by A. A. Jagers (The Netherlands), S. V. Kanetkar, B. Parsons and D. Tyler, and C. Toll.

Pettet noted that it suffices to assume that $[x, \phi(y)] = [\phi(x), y]$ for all x, y; and G need only satisfy $G' = G$ and $Z(G) = 1$.

Convex Functions and Unit Bandwidth

E 3047 [1984, 369]. *Proposed by C. Douglas Harper and Bruce Reznick, University of Illinois at Urbana-Champaign.*

Suppose f is a convex function defined on R and let $U = \{(x, y) : f(x) - 1 \leqslant y \leqslant f(x)\}$. Must U contain line segments of arbitrary length?

Solution by O. P. Lossers, Eindhoven University of Technology, The Netherlands. The answer is 'yes'. For, suppose that f is a convex function, defined on R, such that its parallel strip U of vertical width 1 does not contain a segment of length l, say. Obviously $l > 1$, and we may assume that $f(x)$ is increasing for $x \geq x_0$, $f(x) \to \infty$ as $x \to \infty$. (Also f is continuous.)

Consider a sequence $P_0 P_1 P_2 \ldots$ on the lower boundary of U, $P_i = (x_i, f(x_i) - 1)$ for all i, $x_0 < x_1 < x_2 \cdots$, and $\|P_i P_{i+1}\| = l$ for all i. $P_i P_{i+1}$ contains a point Q_i on the upper boundary of U, so we have a point R_i, one unit below Q_i, on the lower boundary, and we complete the parallelogram $P_i Q_i R_i S_i$. Then it is easily seen that for all i

$$\text{Slope } P_i P_{i+1} < \text{Slope } S_i P_{i+1} < \text{Slope } R_i P_{i+1} < \text{Slope } P_{i+1} P_{i+2}.$$

Now we put $\tan \phi_i =$ Slope $P_i P_{i+1}$ and find

$$\text{Slope } S_i P_{i+1} = \frac{l \sin \phi_i + 1}{l \cos \phi_i} = \tan \phi_i + \frac{1}{l \cos \phi_i}.$$

Hence

$$\tan \phi_i + \frac{1}{l \cos \phi_i} < \tan \phi_{i+1} \quad \text{for all } i,$$

or, by squaring, we see that

$$\frac{1}{\cos^2 \phi_i} - 1 + 2\frac{\sin \phi_i}{l \cos^2 \phi_i} + \frac{1}{l^2 \cos^2 \phi_i} < \frac{1}{\cos^2 \phi_{i+1}} - 1,$$

which implies

$$\cos \phi_{i+1} < \left(\frac{l^2}{l^2 + 1}\right)^{1/2} \cos \phi_i,$$

so that

$$x_N = x_0 + \sum_{i=0}^{N-1} l \cos \phi_i$$

has a limit, L say, as $N \to \infty$, whereas $\lim_{x \uparrow L} f(x) = \infty$, a contradiction. This proves the assertion.

Also solved by F. S. Cater, L. F. Meyers, D. Neuenschwander (student, Switzerland), P. Tracy, and the proposers.

Reznick notes that if f need no longer be defined on all of $\mathbb{R}$, then the answer is "no" even if U is unbounded. For example, let $f(x) = -\log(1 - x)$ for $0 \leq x < 1$, $f''(x) = (1 - x)^{-2} \geq 0$ and consider the tangent line at $x = x_0$. This line intersects $y = f(x) - 1$ when

$$-\log(1 - x) - 1 = -\log(1 - x_0) + (x - x_0)/(1 - x_0),$$

or letting $\alpha = (x - x_0)/(1 - x_0)$, when $1 + \alpha + \log(1 - \alpha) = 0$. This transcendental equation has exactly two solutions $\alpha = \alpha_i$, $\alpha_1 \simeq .8414$, $\alpha_2 \simeq -2.1462$. The line segments extend from $(x_0, f(x_0))$ to $(x_0 + \alpha_i(1 - x_0), f(x_0) + \alpha_i)$ and so have length $|\alpha_i| \cdot (1 + (1 - x_0)^2)^{1/2}$, which is bounded.

$$\boldsymbol{f\left(2x - \frac{f(x)}{m}\right) = mx}$$

E 3053 [1984, 438]. *Proposed by Irl C. Bivens, Davidson College.*

Let m denote a fixed nonzero real number. Find all continuous functions $f: R \to R$ such that

$$f\left(2x - \frac{f(x)}{m}\right) = mx.$$

(Compare with E 2893 [1981, 444; 1982, 702].)

Solution by Víctor Hernández, Universidad Autónoma de Madrid, Spain. The solution is $f(x) = m(x - c)$.

Let $g(x) = 2x - \dfrac{f(x)}{m}$, then the function $g(x)$ is continuous and satisfies

(1) $$g(g(x)) = 2g(x) - x.$$

The function $g(x)$ is one-one and also monotone. If $g(x) = g(x')$, then $g(g(x)) = g(g(x'))$ and by (1) we have $x = x'$. From the relation

$$g(g(x)) - g(x) = g(x) - x$$

we obtain easily that $g(x)$ is increasing. Applying (1) repeatedly we have

$$g^{(n)}(x) = ng(x) - (n - 1)x \quad \text{for each integer } n \geq 1$$

and so

(2) $$g^{(n)}(x) - g^{(n)}(0) = n(g(x) - x - g(0)) + x, \quad n \geq 1.$$

Since g is increasing, letting $n \to +\infty$ it follows that

$$g(x) \leq x + g(0), \quad \text{if } x < 0,$$
$$g(x) \geq x + g(0), \quad \text{if } x > 0.$$

So the range of g is $\mathbb{R}$, and then $g^{(k)}(x)$ is defined for each integer k (i.e., $g^{(-2)}(x) = g^{-1}(g^{-1}(x))$) and the equalities (1) and (2) have sense for every integer n. Again since g is increasing, letting $n \to -\infty$,

$$g(x) \leq x + g(0), \quad \text{if } x > 0,$$
$$g(x) \geq x + g(0), \quad \text{if } x < 0.$$

Then $g(x) = x + c$ and so $f(x) = m(x - c)$.

Also solved by K. Bernstein, U. Everling (West Germany), I. M. Isaacs, B. Klein, O. P. Lossers (The Netherlands), W. Newcomb, N. Passell, University of South Alabama Problem Group, A. Zulauf (New Zealand) and the proposer. Partially solved by W. Janous (Austria).

Two Inequalities for a Triangle

E 3054 [1984, 515]. *Proposed by Vania D. Mascioni (student), Swiss Federal Institute of Technology, Zurich.*

If $\alpha_1, \alpha_2, \alpha_3$ are the angles of a triangle with sides a_1, a_2, a_3, inradius r, area A and

semi-perimeter s, prove the following inequalities:

(a) $$\prod_{i=1}^{3} \alpha_i a_i \geqslant \left(\frac{2\pi}{3}\right)^3 rA,$$

(b) $$\prod_{i=1}^{3} (\pi - \alpha_i) a_i \geqslant \left(\frac{4\pi\sqrt{3}}{9}\right)^3 sA.$$

When does equality hold?

Solution by Klaus Zacharias, Berlin, East Germany. From the Law of Cosines

$$a_i^2 = a_{i+1}^2 + a_{i+2}^2 - 2a_{i+1}a_{i+2}\cos\alpha_i$$

(indices reduced appropriately) follows

$$\cos^2\frac{\alpha_i}{2} = \frac{1}{2}(1 + \cos\alpha_i) = \frac{s(s-a_i)}{a_{i+1}a_{i+2}},$$
$$\sin^2\frac{\alpha_i}{2} = \frac{1}{2}(1 - \cos\alpha_i) = \frac{(s-a_{i+1})(s-a_{i+2})}{a_{i+1}a_{i+2}}.$$

Using Heron's formula $A^2 = s\prod_{i=1}^3(s - a_i)$ and $A = rs$, we obtain

$$\prod_{i=1}^{3}\cos^2\frac{\alpha_i}{2} = s^3\prod_{i=1}^{3}\frac{(s-a_i)}{a_i^2} = s^2A^2\Big/\prod_{i=1}^{3}a_i^2,$$
$$\prod_{i=1}^{3}\sin^2\frac{\alpha_i}{2} = \prod_{i=1}^{3}\left(\frac{s-a_i}{a_i}\right)^2 = A^4\Big/\left(s^2\prod_{i=1}^{3}a_i^2\right).$$

And, since

$$0 < \frac{\alpha_i}{2} < \frac{\pi}{2} \quad\text{and}\quad 0 < \frac{\pi}{2} - \frac{\alpha_i}{2} < \frac{\pi}{2},$$

$$\prod_{i=1}^{3}\alpha_i a_i = 8rA\prod_{i=1}^{3}\frac{\frac{\alpha_i}{2}}{\sin\frac{\alpha_i}{2}}, \quad \prod_{i=1}^{3}(\pi - \alpha_i)a_i = 8sA\prod_{i=1}^{3}\frac{\frac{\pi}{2} - \frac{\alpha_i}{2}}{\cos\frac{\alpha_i}{2}}.$$

So the inequalities are equivalent to

(1) $$\prod_{i=1}^{3}\frac{\frac{\alpha_i}{2}}{\sin\frac{\alpha_i}{2}} \geq \left(\frac{\pi}{3}\right)^3, \quad \prod_{i=1}^{3}\frac{\frac{\pi}{2} - \frac{\alpha_i}{2}}{\cos\frac{\alpha_i}{2}} \geq \left(\frac{2\pi\sqrt{3}}{9}\right)^3.$$

Now we remark that the function

$$f(x) = \ln(x/\sin x)$$

is strictly convex for $0 < x < \pi/2$, because

$$f''(x) = -\frac{1}{x^2} + \frac{1}{\sin^2 x} > 0, \quad x \in (0, \pi/2).$$

(Obviously $x > \sin x > 0$ for $x \in (0, \pi/2)$.) Jensen's inequality

$$\frac{1}{3}(f(x_1) + f(x_2) + f(x_3)) \geq f\left(\frac{x_1 + x_2 + x_3}{3}\right)$$

gives, with $x_i = \alpha_i/2$ ($i = 1,2,3$),

$$\frac{1}{3}\sum_{i=1}^{3}\ln\left(\frac{\frac{\alpha_i}{2}}{\sin\frac{\alpha_i}{2}}\right) \geq \ln\left(\frac{\frac{\pi}{6}}{\sin\frac{\pi}{6}}\right) = \ln\left(\frac{\pi}{3}\right)$$

and, with $x_i = (\pi/2) - (\alpha_i/2)$,

$$\frac{1}{3}\sum_{i=1}^{3}\ln\left(\frac{\frac{\pi}{2}-\frac{\alpha_i}{2}}{\sin\left(\frac{\pi}{2}-\frac{\alpha_i}{2}\right)}\right) \geq \ln\left(\frac{\frac{\pi}{3}}{\sin\frac{\pi}{3}}\right) = \ln\left(\frac{2\pi\sqrt{3}}{9}\right).$$

Taking the exponential proves (1); equality holds if and only if $\alpha_i = \pi/3$ (i.e., for the equilateral triangle).

Also solved by F. F. Abi-Khuzam (Lebanon), S. Arslanagić (Yugoslavia), M. Bencze (Romania), E. Braüne (Austria), W. Janous (Austria), L. Kuipers (Switzerland), P. Kumar (India), W. A. Newcomb, K. L. Stellmacher, M. Vowe (Switzerland), C. Zacharias (India), and the proposer.

Cycles and Transpositions

E 3058 [1984, 516]. *Proposed by Allen J. Schwenk, U. S. Naval Academy.*

(a) We all know that a permutation which is an n-cycle, for example $(1\ 2\ 3 \ldots n)$, cannot be written as a product of fewer than $n-1$ transpositions. Prove it.

(b) In how many ways can $(1\ 2\ 3 \ldots n)$ be written as a product of precisely $n-1$ transpositions?

Solution by O. P. Lossers, Eindhoven University of Technology, The Netherlands.

(a) With every transposition $(i\ j)$, $i, j \in \{1, \ldots, n\}$, we associate an undirected edge joining the two points i and j of a labeled point set $\{1, 2, \ldots, n\}$. In this way every product of k transpositions corresponds uniquely to a graph of k numbered edges on n numbered points. The numbering of its edges is chosen in accordance with the order of the transpositions that are to be carried out in the product. If this product is an n-cycle, each point maps into each other point after a suitable number of iterations. Therefore, this graph must be connected and hence must contain at least $(n-1)$ edges. So $(n-1)$ transpositions have to be involved.

(b) Now let us consider a connected graph with $(n-1)$ edges that contains all of the n points; in other words, we consider a tree. The number of point-labeled trees on n points is equal to n^{n-2} (cf. J. Riordan, *Combinatorial Identities*, p. 168), so the number of point- and edge-labeled trees equals $n^{n-2}(n-1)!$

We shall now show by induction on n that such a tree corresponds to an n-cycle when we consider its transposition product image as pointed out in part (a).

The cases $n = 1$ and $n = 2$ are trivial, because (1) and (12) are cycles. Now let $n > 2$ and assume that all point- and edge-labeled trees on fewer than n points correspond to complete cycles on those points.

Let $(i\ j)$ be the edge corresponding to the first transposition in a product of $n-1$ transpositions. Drop edge $(i\ j)$ but preserve its end points. We get two disjoint trees T and T' on point sets containing k and l points respectively. $k, l > 0$, $k + l = n$, whose transposition product can be written by induction as, say,

$$(i\ a_1\ a_2\ \cdots\ a_{k-1}) \quad \text{and} \quad (j\ b_1\ \cdots\ b_{l-1}).$$

It is important to note that transpositions within T commute with transpositions in T', so that

the final transposition product reads as

$$(j\ b_1\ \cdots\ b_{l-1})(i\ a_1\ \cdots\ a_{k-1})(i\ j) = (i\ b_1\ \cdots\ b_{l-1}\ j\ a_1\ \cdots\ a_{k-1}),\quad \text{an } n\text{-cycle!}$$

Because there are $(n-1)!$ different n-cycles, each n-cycle can be represented as a product of $(n-1)$ transpositions in exactly

$$\frac{n^{n-2}(n-1)!}{(n-1)!} = n^{n-2}$$

ways.

Also solved by G. Behrendt (West Germany), D. M. Bloom, E. D. Bolker, Cal Poly Pomona Problems Group, A. A. Jagers (The Netherlands), S. V. Kanetkar, D. E. Knuth, O. P. Lossers (The Netherlands, second solution), S. Muralidharan (India), D. Neuenschwander (student, Switzerland), F. W. Schmidt and R. Simion, P. Tracy, G. Walls and K. R. Fawcett, P. Y. Wu (Republic of China), and the proposer.

P. Y. Wu pointed out that this problem is contained as a corollary in J. Denes, "The representation of a permutation as the product of a minimal number of transpositions, and its application to the theory of graphs", *Publ. Math. Institute Hung. Acad. Sci.*, 4 (1959) 63-71. G. Behrendt noted, in addition, that part (a) appears as a problem in J. D. Dixon, *Problems in Group Theory*, Dover, New York, 1973. E. Bolker remarked that this problem is also a corollary in E. Bolker and D. Robbins, "The bias of the pseudo-random shuffles", *Aequationes Mathematicae*, 22 (1981) 168-292.

ADVANCED PROBLEMS

For instructions about submitting solutions of these Advanced Problems, which should be mailed by April 30, 1987, see the inside front cover. The solver's full post-office address should be on each sheet.

6531. *Proposed by R. H. Jeurissen, Katholieke Universiteit, Nijmegen, The Netherlands.*

Let G be the (bipartite) graph of which the points are the elements of the symmetric group S_n, with σ and τ adjacent if and only if there is an i, $0 < i < n$, with $\sigma = \tau(i\ i+1)$, i.e., the sequence $\sigma(1), \sigma(2), \ldots, \sigma(n)$ can be transferred into $\tau(1), \tau(2), \ldots, \tau(n)$ by interchanging two consecutive elements. Determine the automorphism group of G (cf. problem 6486 [1985, 62; 1986, 574]).

---

176. MISCELLANEA

The value of research!

I have not needed to think of projective geometries over finite fields since my undergraduate days 20 years ago, and so I was sufficiently rusty for it to take me two evenings to find that the 336 element group underlying Isis Major (RW 14th Feb., p. 131) is the group of symmetries of the projective line over the seven element field.

The discovery that this group can be used, not only for what most of us consider an obscure backwater of geometry, but also for something practical like an eight-bell principle, illustrates the importance of keeping research pure mathematicians adequately funded. No matter how remote from the concrete world their work may seem, these seemingly obscure topics have an uncanny knack of eventually finding useful applications.

–A.G. Smith, from The Ringing World, March 7, 1986.

(The Ringing World is the journal of the British Central Council of Change Ringers.)

6532. *Proposed by Victor Pambuccian, Bucharest, Romania.*

Prove or disprove the following "converse" of Edelstein's theorem (see this MONTHLY, S 8 [1979, 222; 1980, 487]):

Let X be a set of power less than or equal to $2^{\aleph_0}$, and $f\colon X \to X$ a map such that each f^n $(n = 1, 2, \dots)$ has a unique fixed point. Here f^n means $f \circ f \circ \cdots \circ f$ (n times). Then there exists a metric d on X such that (X, d) is a compact metric space and

$$d(f(x), f(y)) < d(x, y) \quad \forall x, y \in X,\ x \neq y.$$

Note that a corresponding "converse" holds for Banach's fixed point theorem ("contraction principle") (see C. Bessaga, Colloq. Math., 7 (1959) 41–43).

SOLUTIONS OF ADVANCED PROBLEMS

Theta Inversion, Heisenberg Uncertainty, and Radar

6491 [1985, 217]. *Proposed by Jon Borwein, Dalhousie University, Canada.*

(a) Show that

$$4\pi = \frac{\sum_{n=-\infty}^{\infty} e^{-n^2\pi}}{\sum_{n=-\infty}^{\infty} n^2 e^{-n^2\pi}}.$$

(b) More generally, show that for each positive integer k

$$4\pi = \frac{\sum_{n=0}^{\infty} k r_k(n) e^{-n\pi}}{\sum_{n=0}^{\infty} n r_k(n) e^{-n\pi}},$$

where $r_k(n)$ is the number of distinct representations of n as a sum of k integral squares.

Solution by William A. Newcomb, Lawrence Livermore National Laboratory. Let S_1, S_2, S_3 and S_4 denote the four series appearing in the statement of the problem, numbered from top to bottom. By the Poisson summation formula,

$$\begin{aligned} S_2 &= \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} t^2 e^{-\pi t^2} \cos 2\pi n t \, dt \\ &= \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} (1 - 2\pi n^2) e^{-\pi n^2} = \frac{1}{2\pi} S_1 - S_2, \end{aligned}$$

so $S_1/S_2 = 4\pi$. Next,

$$\begin{aligned} S_3 &= k \sum_{n=0}^{\infty} e^{-n\pi} \sum_{i_1^2 + \cdots + i_k^2 = n} 1 \\ &= k \sum_{i_1=-\infty}^{\infty} e^{-\pi i_1^2} \cdots \sum_{i_k=-\infty}^{\infty} e^{-\pi i_k^2} = k S_1^k \end{aligned}$$

and

$$S_4 = \sum_{i_1=-\infty}^{\infty} \cdots \sum_{i_k=-\infty}^{\infty} \left(i_1^2 + \cdots + i_k^2\right) e^{-\pi i_1^2} \cdots e^{-\pi i_k^2} = k S_2 S_1^{k-1};$$

the final transposition product reads as

$$(j\ b_1\ \cdots\ b_{l-1})(i\ a_1\ \cdots\ a_{k-1})(i\ j) = (i\ b_1\ \cdots\ b_{l-1}\ j\ a_1\ \cdots\ a_{k-1}),\quad \text{an } n\text{-cycle!}$$

Because there are $(n-1)!$ different n-cycles, each n-cycle can be represented as a product of $(n-1)$ transpositions in exactly

$$\frac{n^{n-2}(n-1)!}{(n-1)!} = n^{n-2}$$

ways.

Also solved by G. Behrendt (West Germany), D. M. Bloom, E. D. Bolker, Cal Poly Pomona Problems Group, A. A. Jagers (The Netherlands), S. V. Kanetkar, D. E. Knuth, O. P. Lossers (The Netherlands, second solution), S. Muralidharan (India), D. Neuenschwander (student, Switzerland), F. W. Schmidt and R. Simion, P. Tracy, G. Walls and K. R. Fawcett, P. Y. Wu (Republic of China), and the proposer.

P. Y. Wu pointed out that this problem is contained as a corollary in J. Denes, "The representation of a permutation as the product of a minimal number of transpositions, and its application to the theory of graphs", *Publ. Math. Institute Hung. Acad. Sci.*, 4 (1959) 63-71. G. Behrendt noted, in addition, that part (a) appears as a problem in J. D. Dixon, *Problems in Group Theory*, Dover, New York, 1973. E. Bolker remarked that this problem is also a corollary in E. Bolker and D. Robbins, "The bias of the pseudo-random shuffles", *Aequationes Mathematicae*, 22 (1981) 168-292.

ADVANCED PROBLEMS

For instructions about submitting solutions of these Advanced Problems, which should be mailed by April 30, 1987, see the inside front cover. The solver's full post-office address should be on each sheet.

6531. *Proposed by R. H. Jeurissen, Katholieke Universiteit, Nijmegen, The Netherlands.*

Let G be the (bipartite) graph of which the points are the elements of the symmetric group S_n, with σ and τ adjacent if and only if there is an i, $0 < i < n$, with $\sigma = \tau(i\ i+1)$, i.e., the sequence $\sigma(1), \sigma(2), \ldots, \sigma(n)$ can be transferred into $\tau(1), \tau(2), \ldots, \tau(n)$ by interchanging two consecutive elements. Determine the automorphism group of G (cf. problem 6486 [1985, 62; 1986, 574]).

---

176. MISCELLANEA

The value of research!

I have not needed to think of projective geometries over finite fields since my undergraduate days 20 years ago, and so I was sufficiently rusty for it to take me two evenings to find that the 336 element group underlying Isis Major (RW 14th Feb., p. 131) is the group of symmetries of the projective line over the seven element field.

The discovery that this group can be used, not only for what most of us consider an obscure backwater of geometry, but also for something practical like an eight-bell principle, illustrates the importance of keeping research pure mathematicians adequately funded. No matter how remote from the concrete world their work may seem, these seemingly obscure topics have an uncanny knack of eventually finding useful applications.

–A.G. Smith, from The Ringing World, March 7, 1986.

(The Ringing World is the journal of the British Central Council of Change Ringers.)

6532. *Proposed by Victor Pambuccian, Bucharest, Romania.*

Prove or disprove the following "converse" of Edelstein's theorem (see this MONTHLY, S 8 [1979, 222; 1980, 487]):

Let X be a set of power less than or equal to $2^{\aleph_0}$, and $f\colon X \to X$ a map such that each f^n $(n = 1, 2, \dots)$ has a unique fixed point. Here f^n means $f \circ f \circ \cdots \circ f$ (n times). Then there exists a metric d on X such that (X, d) is a compact metric space and

$$d(f(x), f(y)) < d(x, y) \quad \forall x, y \in X,\ x \neq y.$$

Note that a corresponding "converse" holds for Banach's fixed point theorem ("contraction principle") (see C. Bessaga, Colloq. Math., 7 (1959) 41–43).

SOLUTIONS OF ADVANCED PROBLEMS

Theta Inversion, Heisenberg Uncertainty, and Radar

6491 [1985, 217]. *Proposed by Jon Borwein, Dalhousie University, Canada.*

(a) Show that

$$4\pi = \frac{\sum_{n=-\infty}^{\infty} e^{-n^2\pi}}{\sum_{n=-\infty}^{\infty} n^2 e^{-n^2\pi}}.$$

(b) More generally, show that for each positive integer k

$$4\pi = \frac{\sum_{n=0}^{\infty} k r_k(n) e^{-n\pi}}{\sum_{n=0}^{\infty} n r_k(n) e^{-n\pi}},$$

where $r_k(n)$ is the number of distinct representations of n as a sum of k integral squares.

Solution by William A. Newcomb, Lawrence Livermore National Laboratory. Let S_1, S_2, S_3 and S_4 denote the four series appearing in the statement of the problem, numbered from top to bottom. By the Poisson summation formula,

$$\begin{aligned} S_2 &= \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} t^2 e^{-\pi t^2} \cos 2\pi n t\, dt \\ &= \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} (1 - 2\pi n^2) e^{-\pi n^2} = \frac{1}{2\pi} S_1 - S_2, \end{aligned}$$

so $S_1/S_2 = 4\pi$. Next,

$$\begin{aligned} S_3 &= k \sum_{n=0}^{\infty} e^{-n\pi} \sum_{i_1^2 + \cdots + i_k^2 = n} 1 \\ &= k \sum_{i_1=-\infty}^{\infty} e^{-\pi i_1^2} \cdots \sum_{i_k=-\infty}^{\infty} e^{-\pi i_k^2} = k S_1^k \end{aligned}$$

and

$$S_4 = \sum_{i_1=-\infty}^{\infty} \cdots \sum_{i_k=-\infty}^{\infty} \left(i_1^2 + \cdots + i_k^2\right) e^{-\pi i_1^2} \cdots e^{-\pi i_k^2} = k S_2 S_1^{k-1};$$

therefore $S_3/S_4 = S_1/S_2 = 4\pi$.

A similar solution is given by Martin Schmidt (West Germany) who begins with the equality of the k-dimensional Gaussian and its Fourier transform, applies the Laplacian operator to both of these, and sums over all k-dimensional lattice points. Newcomb's solution is a shade more direct than that of most readers, who performed a logarithmic differentiation of the theta function inversion formula. L. E. Clarke (England) deduced the result by logarithmic differentiation of Jacobi's triple product identity, together with

$$2\pi \sum_{n=1}^{\infty} \operatorname{sech}^2\left(n - \frac{1}{2}\right)\pi = 1.$$

Part (a) appears explicitly on p. 109 of Walter Schempp, *Radar ambiguity functions, the Heisenberg group, and holomorphic theta series*, Proc. Amer. Math. Soc., 92 (1984), 103-110. It is deduced by taking $m = 1$, $n = 0$ in Schempp's identity

$$\sum_{\mu,\nu=-\infty}^{\infty} L_m\left(\pi(\mu^2+\nu^2)\right) L_n\left(\pi(\mu^2+\nu^2)\right)$$

$$= \frac{n!}{m!}\pi^{m-n} \sum_{\mu,\nu=-\infty}^{\infty} (\mu^2+\nu^2)^{m-n}\left(L_n^{(m-n)}\left(\pi(\mu^2+\nu^2)\right)\right)^2,$$

where $m \geqslant n$ and $L_m^{(\alpha)}(z)$ is the mth Laguerre-Weber function of order α. This in turn is a consequence of a "Poisson-Plancherel identity" (Theorem 8, p. 109) for radially symmetric radar autoambiguity functions (established via harmonic analysis on a compact Heisenberg nilmanifold). This identity is closely connected to the radar uncertainty principle, a relative of the Heisenberg uncertainty principle in quantum mechanics. When radar is used to determine the range and velocity of a target, an increase in the range accuracy entails a decrease in the velocity accuracy and vice-versa. An excellent introduction to uncertainty principles in general is given in H. Dym and H.P. McKean, *Fourier Series and Integrals*, Academic Press, New York, 1972, pp. 116–132.

Martin Schmidt (a thesis student of Schempp's) has also shown that

$$\sum_{n=-\infty}^{\infty} (2\sqrt{\pi}\, n)^{2m} e^{-\pi n^2} = (-1)^n \sum_{n=-\infty}^{\infty} H_{2m}(\sqrt{\pi}\, n)\, e^{-\pi n^2},$$

where H_k is the kth Hermite polynomial. For more on this topic see M. Schmidt, *Die reelle Heisenberg-Gruppe und einige ihrer Anwendungen in Radarortung und Physik*, Diplomarbeit, Universität Siegen, 1985; C. C. Grosjean, *Note on two identities mentioned by Dr. W. Schempp ...*, Proc. Laguerre Symposium, Bar-le-Duc, 1984, Lecture Notes in Mathematics, Berlin, Springer, 1985; and also W. Schempp, *Theta identities via Laguerre functions*, 1984 (preprint).

Also solved by 17 other readers and the proposer.

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER

The Genesis of the Abstract Group Concept: A Contribution to the History of the Origin of Abstract Group Theory. By Hans Wussing. Translated by Abe Shenitzer. The MIT Press, Cambridge,

MA, 1984. 331 pp. $30.00 (cloth).

KAREN HUNGER PARSHALL
Department of Mathematical Sciences, Sweet Briar College, Sweet Briar, VA 24595

How did the mathematical notion of a group arise? The standard textbook presentation of group theory suggests the following historical development. One day, it is not clear when, a mathematician sat down and decided to invent a new mathematical object. Starting with a set and a binary operation on its elements, the mathematician wrote down three rules. For the set to be a group, for that was the name the inventor chose for the creation, it must satisfy each of the rules relative to the operation. What could be proven about a group so defined? A few facts were obvious and quickly shown, but they were easy and not terribly challenging. The mathematician thought a bit more about the new group and decided to define a subgroup, a collection of elements in the original group which also satisfied the three laws. This was much more interesting. Things seemed to fit together perfectly, and more theorems resulted. By this time word of the new invention had spread and other mathematicians decided to try their skills at groups. They made new definitions and proved new results (some even had theorems named for themselves), and the new creation generated an entire discipline, which became known as the theory of groups.

Authors of modern algebra textbooks are not really to blame for implying this warped historical perspective of group theory. After all, their task is to present new material in the most pedagogically expedient way possible. The axiomatic approach, which starts with a devilishly simple definition and logically proceeds to results of ever-increasing complexity, provides a neat and clean, if somewhat sterile, look at the subject. In actuality, though, the theory of groups has evolved through a complicated and often surprising interconnection of subjects and ideas with axiomatization coming only very late in the process.

In his book, *The Genesis of the Abstract Group Concept*: *A Contribution to the History of the Origin of Abstract Group Theory*, originally appearing in German in 1969 and now appearing for the first time in English translation, Hans Wussing traced the evolution of the group concept from its disparate sources in the eighteenth and early nineteenth centuries through its full-blown emergence at the beginning of the twentieth century. The historical development he uncovered differed markedly from the traditional, and as it turned out, simplistic account found, for example, in Eric Bell's *The Development of Mathematics* or Dirk Struik's *A Concise History of Mathematics*.[1] In these treatments, the theory of algebraic equations spawned the concept of a group of permutations, which generalized to the abstract notion of a group.[2] Wussing's research reveals a much more complicated situation.

The group concept grew not from one but from three distinct yet intertwined roots: the traditionally recognized theory of algebraic equations, number theory, and geometry. Furthermore, each of these roots shared equally in feeding the growing idea. As Wussing explains in his introduction: "The existence of two additional roots of abstract group theory has been obscured mainly by the fact that the group-theoretic modes of thought in number theory and geometry remained implicit until the end of the middle third of the nineteenth century; they made no use of the term 'group' and, in the beginning, had virtually no link to the contemporary development of the theory of permutation groups" (p. 17). Wussing's self-proclaimed task, then, was to locate "... those paths of development of implicit group theory that have made a causal contribution to the rise of explicit group theory" (p. 17). This may sound like so much double-talk, but it really is not. Wussing realized that tracing the development of the group concept involves more than searching the literature for occurrences of the word "group". Patterns of thought and combinations of properties cropped up over and over again long before they became universally distinguished by a special nomenclature. In his study Wussing pursued these much more elusive phenomena and provides convincing accounts of their rôles in the development of the group concept.

Aside from the merits of its contents, Wussing's book mirrors two important historiographical trends in the history of modern mathematics. First, his research hinges upon a close and painstaking analysis of the primary sources. He details not only the internal logic of the original research works but also the deep relationships between them. Unlike older and more traditional authors such as Bell and Struik, Wussing is willing to deal with recent mathematics technically in his historical discussion. Second, his findings discount the standard, linear history of the group concept and reveal a tangled web of clashing and competing ideas deriving from various mathematical realms. In virtually all mathematical publications, whether in textbooks or in the most high-powered research articles, mathematical development seems synonymous with cool and calculated logical thought, but its history is really as convoluted and open to interpretation as, for instance, the history of the French Revolution. Wussing's book reflects this more sophisticated view of the history of mathematics and so helps to set a higher standard for historical research.

Fifteen years have now passed since *The Genesis of the Abstract Group Concept* first appeared in German, and in this time standards have continued to change. Appearing in English translation only now, Wussing's book unfortunately seems a bit like a historical document itself. As the author states in his preface to the American edition, the conclusions he reached while writing the book from 1964 to 1966 have largely withstood the proverbial test of time, that is, they have been upheld by unpublished manuscript sources which have come to light in the intervening years. As his extensive (almost 750 entries) bibliography shows, Wussing based his findings almost exclusively on published papers. Today, although published documents obviously retain their importance, historians of mathematics increasingly seek out unpublished sources such as letters, notes, and rough drafts in order to deepen their insights not only into the mathematical ideas but also into the personalities of the mathematicians themselves. Mathematics is a human endeavor. Living, breathing individuals who read newspapers, talk to their friends, take philosophical stances, and maybe even write music or poetry, create mathematics. More and more, historians strive to capture this vitality by blending the technical development of mathematical ideas with the detailed discussion of biographical information. Wussing acknowledges the strictly internalistic nature of his study and admits that a complete history would have more effectively linked the ideas to the people who generated them. His narrative as well as his historical re-creation would have profited from this added dimension.

If Wussing's book falls somewhat short of the current standards of historical scholarship, it is only because it contributed to setting and directing those standards fifteen years ago. Although much has been accomplished in the history of nineteenth century algebra since 1969, and especially in the history of Lie theory,[3] Wussing's interpretation of the development of the group concept has only been extended not surpassed. Any student, young or old, who has been bedazzled by the slick textbook treatment of group theory should find Wussing's *The Genesis of the Abstract Group Concept* eye-opening.

Notes

[1]Eric Temple Bell, *The Development of Mathematics*, 2nd ed. (New York: McGraw-Hill Book Company, 1945) and Dirk J. Struik, *A Concise History of Mathematics*, 3rd rev. ed. (New York: Dover Publications, 1967).

[2]See Bell, pp. 232–233, and Struik, p. 151.

[3]Many of Lie's books as well as his collected works have been reprinted. In particular, Michael Ackerman produced English translations with technical commentary by Robert Hermann of two of Lie's seminal papers. See *Sophus Lie's 1880 Transformation Group Paper*. Translated by Michael Ackerman with Comments by Robert Hermann (Brookline, MA: Math. Sci. Press, 1975) and *Sophus Lie's 1884 Differential Invariant Paper*. Translated by Michael Ackerman with Comments and Additional Material by Robert Hermann (Brookline, MA: Math. Sci. Press, 1976). Furthermore, Thomas Hawkins has been systematically researching various aspects of the history of Lie theory. Among other papers, see his "Hypercomplex Numbers, Lie Groups, and the Creation of Group Representation Theory," *Archive for History of Exact Sciences* 8 (April 1972):243–287, "Non-Euclidean Geometry

and Weierstrassian Mathematics: The Background to Killing's Work on Lie Algebras," *Historia Mathematica* 7 (1980):289–342, and "Wilhelm Killing and the Structure of Lie Algebras," *Archive for History of Exact Sciences* 26 (1982):127–192.

LETTERS TO THE EDITOR

For instructions about submitting letters for publication in this department see the inside front cover.

Editor:

Professor Lehmer's derivation [**1**, Theorem (p. 452)] of a power series for

$$(2x \arcsin x)(1-x^2)^{-1/2}$$

makes use of Gregory's series for arctan t, the binomial theorem (twice), inversion of a double sum, Wallis' formula for $\int_0^{\pi/2} \sin^{2r-1}\theta \, d\theta$, and two changes of variable. I offer the following standard direct method which does not assume the evaluation of any other series or integrals.

Let $y(x) = (\arcsin x)^2$. Differentiation and squaring gives

$$(1-x^2)y'^2 - 4y = 0.$$

A further differentiation shows that $y(x)$ is the solution of the system

$$(1-x^2)y'' - xy' - 2 = 0, \quad y(0) = y'(0) = 0.$$

Each $x \in (-1,1)$ is an ordinary point of this differential equation and so there is a power series $y(x) = \Sigma_0^\infty a_n x^n$ valid for $|x| < 1$ (and the boundary conditions give $a_0 = a_1 = 0$). Term by term differentiation and substitution in the differential equation gives

$$(1-x^2)\sum_0^\infty n(n-1)a_n x^{n-2} - \sum_0^\infty n a_n x^n - 2 = 0,$$

whence $2a_2 = 2$ and $(n+2)(n+1)a_{n+2} = n^2 a_n (n \geqslant 1)$.

Thus for $m = 1, 2, \ldots$ we have $a_{2m-1} = 0$ and

$$a_{2m} = \frac{[2\cdot 4 \cdots (2m-2)]^2}{3\cdot 4\cdot 5 \cdots (2m)} = \frac{2[2^m m!]^2}{(2m)^2(2m)!} = \frac{2^{2m-1}}{m^2\binom{2m}{m}}.$$

Hence

$$2(\arcsin x)^2 = \sum_{m=1}^\infty (2x)^{2m} \Big/ \left[m^2\binom{2m}{m}\right] \quad \text{for } |x|<1 \quad [\text{see } \mathbf{1}\,(13)].$$

Differentiation gives the required series for $(2x \arcsin x)(1-x^2)^{-1/2}$.

Reference

1. D. H. Lehmer, Interesting series involving the central binomial coefficient, this MONTHLY, 92 (1985) 449–457.

Dennis C. Russell
Department of Mathematics
York University
Toronto, Canada M3J 1P3.

U S Postal Service

STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION
Required by 39 U S C 3685)

1A TITLE OF PUBLICATION: THE AMERICAN MATHEMATICAL MONTHLY

1B PUBLICATION NO: A M M Y A E

2 DATE OF FILING: Sept. 30, 1986

3 FREQUENCY OF ISSUE: monthly except bi-monthly June/July and Aug/Sept

3A NO. OF ISSUES PUBLISHED ANNUALLY: ten

3B ANNUAL SUBSCRIPTION PRICE: Library $70.00; Indiv.Member $24.00

4 COMPLETE MAILING ADDRESS OF KNOWN OFFICE OF PUBLICATION *(Street, City, County, State and ZIP+4 Code) (Not printers)*

1529 Eighteenth Street, N.W., Washington, D.C. 20036

5 COMPLETE MAILING ADDRESS OF THE HEADQUARTERS OF GENERAL BUSINESS OFFICES OF THE PUBLISHER *(Not printer)*

1529 Eighteenth Street, N.W., Washington, D.C. 20036

6 FULL NAMES AND COMPLETE MAILING ADDRESS OF PUBLISHER, EDITOR, AND MANAGING EDITOR *(This item MUST NOT be blank)*

PUBLISHER *(Name and Complete Mailing Address)*

The Mathematical Association of America, 1529 Eighteenth St.,N.W., Washington, D.C. 20036

EDITOR *(Name and Complete Mailing Address)*

P.R. Halmos, Univ. of Santa Clara, Dept. of Mathematics, Santa Clara, CA 95053

MANAGING EDITOR *(Name and Complete Mailing Address)*

Dr. Raoul Hailpern, 63 Garden Court, Amherst, NY 14226

7 OWNER *(If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual must be given. If the publication is published by a nonprofit organization, its name and address must be stated.) (Item must be completed.)*

FULL NAME	COMPLETE MAILING ADDRESS
The Mathematical Association of America, Inc.	1529 Eighteenth St., NW Washington, DC 20036

8 KNOWN BONDHOLDERS, MORTGAGEES, AND OTHER SECURITY HOLDERS OWNING OR HOLDING 1 PERCENT OR MORE OF TOTAL AMOUNT OF BONDS, MORTGAGES OR OTHER SECURITIES *(If there are none, so state)*

FULL NAME	COMPLETE MAILING ADDRESS
None	

9 FOR COMPLETION BY NONPROFIT ORGANIZATIONS AUTHORIZED TO MAIL AT SPECIAL RATES *(Section 423.12 DMM only)*
The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes *(Check one)*

(1) [X] HAS NOT CHANGED DURING PRECEDING 12 MONTHS

(2) [ ] HAS CHANGED DURING PRECEDING 12 MONTHS

(If changed, publisher must submit explanation of change with this statement.)

10 EXTENT AND NATURE OF CIRCULATION *(See instructions on reverse side)*	AVERAGE NO. COPIES EACH ISSUE DURING PRECEDING 12 MONTHS	ACTUAL NO. COPIES OF SINGLE ISSUE PUBLISHED NEAREST TO FILING DATE
A TOTAL NO. COPIES *(Net Press Run)*	21,000	21,772
B PAID AND/OR REQUESTED CIRCULATION 1 Sales through dealers and carriers, street vendors and counter sales	100	100
2 Mail Subscription *(Paid and/or requested)*	17,999	18,794
C TOTAL PAID AND/OR REQUESTED CIRCULATION *(Sum of 10B1 and 10B2)*	18,099	18,894
D FREE DISTRIBUTION BY MAIL, CARRIER OR OTHER MEANS SAMPLES, COMPLIMENTARY, AND OTHER FREE COPIES	54	55
E TOTAL DISTRIBUTION *(Sum of C and D)*	18,153	18,949
F COPIES NOT DISTRIBUTED 1 Office use, left over, unaccounted, spoiled after printing	2,847	2,823
2 Return from News Agents	0	0
G TOTAL *(Sum of E, F1 and 2—should equal net press run shown in A)*	21,000	21,772

11 I certify that the statements made by me above are correct and complete

SIGNATURE AND TITLE OF EDITOR, PUBLISHER, BUSINESS MANAGER, OR OWNER: Executive Director, MAA

PS Form 3526, Dec 1985 *(See instruction on reverse)*

JOIN MAA TODAY!

Send this application to: MAA Membership Department
1529 18th Street, N.W.
Washington, D.C. 20036

Name ______________________________

Mailing Address ______________________________

______________________________ Zip ______________

Employer/School ______________________________

Position (Rank) ______________________________

Employer's City/State ______________________________

Highest Earned Degree ______________ Year Degree Earned ______________

Institution Awarding Degree ______________________________

Month/Year of Birth ______________________________

Have you been a member of the MAA before? ☐ **Yes** ☐ **No**

Payment Enclosed $__________ *U.S. Funds Only* (see below)

Your membership may begin on July 1 or January 1. You may elect to receive any of the listed combinations of the three MAA journals:

THE AMERICAN MATHEMATICAL MONTHLY **(M)**
MATHEMATICS MAGAZINE **(G)**
THE COLLEGE MATHEMATICS JOURNAL **(J)**

Please circle the appropriate box corresponding to your initial membership period and the selection of journals you wish to receive. (Rates guaranteed for the indicated periods only.) Subscription prices are included with dues (see **).

	M	**G**	**J**	**M+G**	**M+J**	**G+J**	**M+G+J**
Student Membership*							
1 year (Jan.-Dec. 1986)	$24	$18	$20	$30	$32	$26	$38
1½ year (July 1986-Dec. 1987)	$36	$27	$30	$45	$48	$39	$57
Regular Membership							
1 year (Jan.-Dec. 1986)	$40	$28	$31	$52	$55	$43	$67
1½ year (July 1986-Dec. 1987)	$60	$42	$46.50	$78	$82.50	$64.50	$100.50

*Student Membership available to high school and undergraduate students and to students regularly enrolled in graduate study at least half time. Enclose letter from school or department official confirming status. Student rates apply to unemployed persons under certain circumstances.

**Annual dues include subscription prices as follows: Reg. Memb. $24 (M), $15 (J), $12 (G), $1 (Focus). Std. Memb. $12 (M), $8 (J), $6 (G), $1 (Focus).

METHOD OF PAYMENT ☐ Check enclosed ☐ VISA ☐ MasterCard ☐ Bill me

Card Number |_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_| Expiration Date |_|_|_|_|
(month/year)

Interbank Number |_|_|_|_| *(MasterCard only—located above name on card)*

SIGNATURE ______________________________

Carus Mathematical Monograph #21

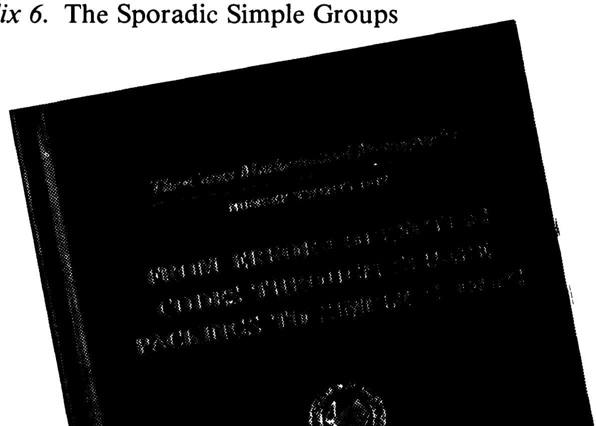
From Error Correcting Codes through Sphere Packings to Simple Groups,

by Thomas M. Thompson
224 pp. Hardbound
List: $24.00 MAA Member: $18.50

Two of the most fascinating problems to challenge mathematicians in recent years concern the construction of data transmission codes that can correct errors introduced by static and the search for efficient ways to pack ping-pong balls into a box. Can one design the best error-correcting codes? Can one find the most efficient sphere packing?

Therein lies a fascinating story which is told with great skill and clarity in this important addition to the Carus Mathematical Monograph series. The author has packed (sic) into 175 pages all of the basic mathematical ideas of this saga woven into a gripping historical account of the journey from error-correcting codes to sphere packings to simple groups.

Table of Contents



THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.